SHERPA

Shaping the ethical dimensions of smart information systems– a European perspective (SHERPA)

Deliverable No. 1.2

SIS Scenarios

30 April 2019

Disclaimer: The views expressed in this deliverable are those of the authors of the scenarios and those who contributed to the scenarios. In no way, do they reflect the views of the European Commission.

**This project has received funding from the
European Union's Horizon 2020 Research and Innovation Programme
Under Grant Agreement no. 786641**

Document Control

Deliverable	D1.2 – SIS Scenarios
WP/Task Related	WP1 - Representation and visualization of ethical and human rights issues in SIS. Task 1.2 – Develop five SIS scenarios
Delivery Date	30 April 2019
Dissemination Level	PU
Lead Partner	David Wright (TRI)
Contributors	Rowena Rodrigues (TRI), Tally Hatzakis (TRI), Corinna Pannofino (TRI), Kevin McNish (Twente), Mark Ryan (Twente), Bernd Stahl (DMU) Josephina Antoniu (UCLAN)
Reviewers	Doris Schroeder (UCLAN), workshop participants, other stakeholders
Abstract	This deliverable considers the ethical, legal (data protection), social and economic impacts of new and emerging technologies, powered by artificial intelligence (AI) and big data, which we call smart information systems (SIS). In this deliverable, we look forward to the year 2025 to consider how new and emerging technologies may raise various issues, regarding which **policymakers and other stakeholders should consider what ethical guidelines, data protection policies and other measures we might need to address the issues now rather than five or six years from now when they may have fewer policy options.** We have developed five scenarios, addressing five different technology clusters in five different areas – social care for senior citizens, information warfare, predictive policing, driverless cars and learning buddy robots. We engaged stakeholders in the development of these scenarios. Stakeholder engagement was an important purpose of the scenario construction process. In short, the scenarios are a way of exploring with stakeholders the issues raised by these advanced new technologies and developing recommendations for policymakers for dealing with those issues. This deliverable comprises this executive summary, an introduction, the five scenarios, and conclusions and recommendations.
Key Words	Scenarios, ethics, privacy, emerging technologies, AI, SIS, big data

Revision History

Version	Date	Author(s)	Reviewer(s)	Notes
0.1	30 Jan 2019	David Wright (TRI)	Doris Schroeder, workshop participants, other stakeholders	First Draft
0.2	11 Mar 2019	David Wright		Second Draft

Contents

EXECUTIVE SUMMARY 4

List of figures 14

List of tables 14

List of acronyms/abbreviations 15

GLOSSARY OF TERMS 15

1. OBJECTIVES OF THIS DELIVERABLE 17

2. INTRODUCTION 17

3. FIRST SCENARIO: CREATING COMPANIONS FOR SENIOR CITIZENS WITH TECHNOLOGIES THAT MIMIC PEOPLE 28

4. SECOND SCENARIO: INFORMATION WARFARE IN 2025 44

5. THIRD SCENARIO: PREDICTIVE POLICING IN 2025 62

6. FOURTH SCENARIO: SELF-DRIVING VEHICLES: NAVIGATING TOWARDS AN ETHICAL FUTURE 75

7. FIFTH SCENARIO: ARTIFICIAL INTELLIGENCE & ROBOTS IN EDUCATION IN 2025 99

8. CONCLUSIONS AND RECOMMENDATIONS 116

8.1 The SHERPA scenario construction process 116

8.2 Key conclusions 117

8.3 The bottom line - recommendations 118

Executive Summary

Purpose of the deliverable

This deliverable considers the ethical, legal (data protection), social and economic impacts of new and emerging technologies, powered by artificial intelligence (AI) and big data, which we call smart information systems (SIS). In this deliverable, we look forward to the year 2025 to consider how new and emerging technologies may raise various issues, regarding which **policymakers and other stakeholders should consider what ethical guidelines, data protection policies and other measures we might need to address the issues now rather than five or six years from now when they may have fewer policy options.** We have developed five scenarios, addressing five different technology clusters in five different areas – social care for senior citizens, information warfare, predictive policing, driverless cars and learning buddy robots. We engaged stakeholders in the development of these scenarios. Stakeholder engagement was an important purpose of the scenario construction process. In short, the scenarios are a way of exploring with stakeholders the issues raised by these advanced new technologies and developing recommendations for policymakers for dealing with those issues.

Scope of the deliverable

This deliverable comprises this executive summary, an introduction, the five scenarios, and conclusions and recommendations.

Each of the five scenarios is structured similarly (but not exactly the same – we were not unduly procrustean in the construction and structure of the five scenarios). Each scenario introduces the technologies and applications that may be available in 2025, a brief vignette to illustrate how the technologies or applications may be used, the drivers of those technologies and applications, i.e., the factors that have impelled the development of those technologies, the barriers or impediments to the use of such technologies, the ethical, legal, social and economic impacts, the recommendations to reach a desired future and avoid an undesired future.

We have placed each scenario in a table, with the text of the scenario in the left column. In the right column are a few questions relating to each section of the scenario to help stimulate responses from stakeholders.

Task description

This deliverable has its origin in Task 1.2 of the SHERPA Description of Action (DoA), which forms part of the consortium's Grant Agreement with the European Commission. The task specifies that "SHERPA will develop five scenarios exploring emerging SIS that are likely to be implemented and socially relevant five years hence". The task description says that the scenarios will highlight "critical ethical and human rights issues arising from the use of future SIS. The scenarios will also highlight cyber-security risks." Task 1.2 describes the steps for building the scenarios, which remain as valid after construction of the scenarios as they were before undertaking our effort. Most importantly, Task 1.2 states that "The partners will engage stakeholders in the construction and validation of the scenarios" and that the partners will "disseminate the scenarios and the scenario methodology to a range of stakeholders asking for their views on minimising the risks to ethics and human rights while maximising the economic and societal benefits of the new technologies". The partners have followed the plan laid out in the Task 1.2 description.

Process in creating the scenarios and stakeholder engagement

Our scenario construction methodology engages stakeholders from the start of the process, i.e., the scenario leader organises a kick-off workshop of stakeholders who brainstorm on what 2025 might be like, in

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

particular, in regard to the AI-driven technologies that drive each of the five topic areas. The scenario leader uses the results of the brainstorming session as raw material for constructing the first draft of the scenario. He or she then sends that draft to the workshop participants and ask for their comments. The scenario leader draws on those comments for a second iteration of the scenario, which he or she then sends to a larger group of stakeholders for their comments. In the SHERPA project, this larger group was the project's stakeholder board, which comprises 30 stakeholders. With the comments from the larger group, the scenario leader revises the scenario again and creates a third iteration, which he or she then sends to the project's contact list. Taking into comments received from the contact list, the scenario leader creates a fourth iteration, which he or she posts on the project website and invites comments from visitors.

Why do we go through so many iterations with stakeholders? We operate on the assumption that the more stakeholders who comment on the scenario, the greater the credibility of the resulting scenario, the greater the buy-in from stakeholders, the more legitimate are the recommendations we present to policymakers, particularly those at European level.

The workshops

The first scenario brainstorming workshop, concerning technologies that mimic people, was held in the premises of Innovate UK in Brussels on 3 July 2018. This workshop had 22 participants, most of whom were SHERPA partners (13). In addition, were four stakeholder board members, two EC policy officers and one external stakeholder. Of participants, 12 were women and 10 men. As this was the first time most partners were involved in scenario construction, we focussed the workshop on SHERPA partners and introduced them to the particular scenario methodology developed by Trilateral for the project.

The second scenario workshop, concerning artificial intelligence in warfare, was also held at the premises of Innovate UK in Brussels in the afternoon of 17 September 2018. There were only nine participants in this workshop, including the project officer. All but two were male. Nevertheless, the workshop generated lots of useful discussion, including a presentation by defence journalist Nick Cook that led to the travails of information warfare.

The third scenario workshop, concerning AI in education, was held at the same premises the following morning, with 17 participants, primarily from academia, but also five from partners and one journalist. The gender split was almost even, with nine females and eight males.

Our University of Twente partner hosted the fourth and fifth scenario workshops, which concerned predictive policing and driverless cars (Self-driving vehicles, SDVs) in the afternoon of 25 September and in the morning of 26 September 2018 respectively.

The SDVs workshop had 20 participants from a wide range of backgrounds, experiences and disciplines from academia, the public sector and the private sector, with a 70% - 30% male-female ratio in the group. The partners gave careful attention to ensuring a diversity of approaches and viewpoints were represented in the workshop, with individuals from standardisation bodies; SDV testing; computer scientists; engineers; psychologists; AI specialists; cybersecurity experts; ethicists; and legal scholars. The workshop was split into subgroups for more inclusive and conducive discussion for the construction of the scenario later. These sections were split between group-work, open discussion, and critical dialogue of SDVs.

The final workshop on predictive policing followed the same format in splitting the group into subgroups to address different points on the agenda followed by a plenary group discussion of the points raised in the subgroups.

Summary of each of the five scenarios

The following paragraphs provides a brief summary of each of the scenarios.

The first scenario concerns technologies that mimic people. In the year 2025, such technologies are becoming commonplace. With an ageing population, European governments are finding it increasingly challenging to provide social services and assisted living facilities to all those in need. The situation is becoming harder for those whose partner dies. The following scenario has been designed around this vignette.

Alfred's wife of 45 years died in 2024. He missed her greatly and doctors were worried about his mental health until a government agency told him that he could have a hologram of his dear wife Lucy who could interact with him. The hologram knows about their lives together. Thanks to artificial intelligence, the hologram technology has synthesised all of Lucy's data from her social media and is able to reproduce her voice, her appearance, her mannerisms, even the way she used to laugh with him. Data from their electronic home assistants, Siri and Alexa, were really useful too. The new Lucy reminded Alfred to take his daily medication and to go for a walk because he needed the exercise. Although the research is still preliminary, sociologists and physicians are in general agreement that senior citizens who engage with holograms or personalised avatars are likely to live healthily longer. A public consultation in 2024 showed that a majority of respondents favoured the deployment of holographic support services from 2025 onwards.

Their deployment remains, however, controversial. Some government agencies insist on taking partial control of Lucy and her peers. This is for Alfred's safety and well-being to make sure Lucy functions properly and caters for him, for example, to prompt Alfred to take his medicines and encourage him to do some physical exercise or converse with her instead of watching TV all day long. But activists are suspicious of some lines of the hologram's questioning, for example, why Lucy quizzes Alfred about whether he is working part-time or has any other sources of income, suggesting that governments have an ulterior motivation to reduce his benefits. Privacy advocates have repeatedly expressed concerns that the holograms, avatars or care robots are actually sophisticated surveillance agents as they could pass on the information they collect about their owners to the big tech companies and government agencies. There have also been concerns about whether holograms, like Lucy, can make medical diagnoses. Studies have shown that holographic people are more often right in their diagnoses in 2025 than real doctors.[1] Public opinion is divided.

The second scenario concerns information warfare. It notes that the nature of warfare has changed and so have the instruments of warfare and even the soldiers. Until the advent of the Internet, combatants in warfare were generally states or their proxies. The instruments of warfare – weapons with projectiles – were well known before the Internet, but ineffective in cyberspace. The pre-Internet soldiers were trained and wore uniforms. Today's warrior could take down an energy grid from her bedroom – without firing a shot – simply by pressing some keys on her laptop. If a bomber from one country dropped a bomb on another's country nuclear power plant, it would likely provoke an outbreak of physical war. But in cyber space, a country can disable the power plant with little fear of physical retaliation. The enemy state can deny having been the source of the attack.

In 2025, many states are engaged in information warfare. They have been joined by other "enemy combatants", including criminal gangs, terrorists, rich people with an agenda, political parties, rogue employees, etc., many of whom have become very skilful in covering their digital tracks. It becomes increasingly difficult to unravel what little forensic evidence exists.

In the cyber age, the nature of attacks has changed too. Information warfare takes many more forms than simply disabling a power plant. Today's attackers are disrupting and undermining political processes, to persuade people by purveying fake news and calling into question legitimate news, by gathering huge amounts of personal data of all kinds on whole populations, by holding utilities and vital services, such as the national health services to ransom, by espionage and the theft of all kinds of intellectual property.

[1] See, for example, Whyte, Chelsea, "AI can diagnose childhood illnesses better than some doctors", *New Scientist*, 11 Feb 2019.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Attackers can use AI to manufacture videos of events that never took place or of politicians recorded as saying things they never said. The vignette tells the story of a digital attack on a nuclear power plant. Public opinion presses for retaliation, but it is not clear who is responsible – at least, not yet. Defenders have made advances too – they have algorithmic hunters that search the Internet, including the dark net, for traces of terrorists and criminals. As cyber attackers are increasingly relying on artificial intelligence for unleashing bots, so defenders depend on AI to detect and fend off attacks. Human decision-making is not fast enough, hence, defenders are relying on AI-powered, real-time, automated decision-making to defend their assets in both cyber space and on the ground.

The third scenario concerns predictive policing. In 2025, the financial crisis has meant that the police have had to do more with less. In most Chinese cities, facial recognition on CCTV is now standard. With other means of societal surveillance, such as biometrics on public transport and the universal Social Credit System that tracks bank records and voice recognition ATMs, many see China as having become the archetypal "surveillance state". There have been governmental intrusions upon the privacy of some citizens, but this is a price the majority appear willing to pay for their convenience and safety. The US has been using algorithmic-based predictive policing for some years now; it has ceased to be a "live" political issue. In less wealthy countries, predictive policing systems are used primarily to protect the rich from the poor, creating virtual gated communities. Europe is caught in the middle. The ageing European population fears that the waves of millions of undocumented immigrants coming into Europe will increase crime in Europe. Younger Europeans are more empathetic. Politicians have difficulty developing a consistent and effective response to the immigration issue as well as rising crime. Support for the far right and the far left continues to rise, causing significant societal conflict; each side is compounding social divisions. Violence, fraud, online scams and hacking are all significant problems for social stability.

In response to these challenges, the police need to remain effective and accountable. Smart policing systems that predict the location and sometimes the perpetrators of crimes can help to compensate for the lack of resources. However, they are also criticised for invading the privacy of citizens, and Europe has always seen itself as the voice of reason on human rights. The *European Charter for Fundamental Rights* and the *European Convention on Human Rights* are known globally and are often used as yardsticks on such matters by the UN. Can Europe be seen to backtrack? Can all of these developments be subsumed under the legal exceptions of radicalisation and counter-terrorism, even when many of these approaches are clearly a response to low-level and white-collar crime? Is it time to expand such exceptions to include the promotion of civil unrest? There are some technical solutions: AI that is transparent in its processes, for example, will more likely avoid biases and might help with adherence to fundamental rights, but more needs to be done to restore trust in the police.

The fourth scenario concerns self-driving vehicles (SDVs). In the year 2025, self-driving vehicles can be used in most European cities. Over the past few years, technology has come a long way and safety levels have been consistently better than human driver error for some time. Those who can afford SDVs are usually the cool, trendy, tech types who use the SDV as an extension of the personal work/life space where one can work, sleep, read, eat, watch movies or TV, or just observe their surroundings. Non-owners can rent the machines at charges competitive with those for cars that require a driver. There are few parked cars now, and much less traffic than a decade ago. Public spaces have become more open and you rarely see elderly or disabled people or people with prams struggling to cross streets thronged with cars.

This suits Adrian's lifestyle a lot: "I am able to work in my car, while commuting. When you factor in an hour commute each way, I get back 10 hours of my life from the commute every week. I sit back with my laptop, while listening to Spotify. It's great!" Adrian's Waymo Centauri b is one of the few permitted self-driving car models on the market and has been one of the most widely adopted of these vehicles, so far. Inter-city driving, however, is still "a nuisance when I have to drive outside München. It takes a while to get used to the wheel again," Adrian claims.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Some controversial incidents have made headlines in the past few years. A woman died in labour a few months ago, as the SDV would not exceed the speed limit to rush her to hospital. Her husband did as much as possible under the guidance of doctors via cameras, but it was not enough. Many questioned why he didn't take over the car, but he did not have a driving licence, and, in any event, the car-sharing company would not allow him to take over for insurance reasons. He claimed that if cyber-criminals had taken over the vehicle to create havoc, the car-sharing company or the police would have done something to control the car, so why couldn't they have done something to save his wife? The husband has taken both the car-sharing company and the police to court; the case is ongoing. It often comes up in conversations. Some, especially taxi drivers, are quick to judge their car-sharing competitors harshly. Others point to this being a rare, tragic event and claim one should not overlook all the benefits SDVs have brought, especially since they became electric. Living in cities would be next to impossible without them nowadays.

The fifth scenario concerns learning buddies and other educational technologies. Technological changes have revolutionised the classroom and curricula across schools in Europe. In 2025, artificial intelligence has significantly changed education at all levels. Changes implemented in schools are now moving outside the classroom. Companies are advertising new employment opportunities, tailored to the new curricula. Currently, the changes in education in the last seven years can be categorised into four types:

i. *Move towards collaborative learning*
ii. *Use of automation to provide assessment feedback*
iii. *Personalisation using big data*
iv. *Visualisation that allows visiting extraordinary scenarios* (AI-powered virtual reality and augmented reality).

The scenario envisages a future with learning buddies, robots that mentor their young charges, at their command with huge amounts of data from many different sources at their disposal. Educational technologies are redefining the role of the teacher to become a *facilitator* of the learning activity. Because information is becoming ubiquitous, teachers employ technology to help students in improving their reasoning and critical thinking skills. AI-powered robots help overworked professors to answer thousands of questions over the course of a semester. The public is conflicted on the use of robots. Some people fear that big data means intrusions upon their privacy, which is the central ethical issue, since there will be a generation, collection and manipulation of personal data, specifically of sensitive data. The classroom robots and learning buddies are constantly collecting data from their environment, including interacting with the students, via video and audio monitoring (surveillance) of the humans.

Main recommendations

In the following paragraphs, we present some, not all, of the key recommendations made in each scenario. All of the recommendations can be found at the conclusion of each scenario.

Holographic companions for senior citizens

- Academics should explore the ramifications of the new technologies and, where possible, ensure technologies that mimic people are open source. Before holograms or robots are used in social care applications, such as that depicted in the vignette, developers and/or operators should conduct a data protection impact assessment.
- Industry should develop and use ethics councils within individual companies and as well as across companies. Such councils should be multi-disciplinary with people from backgrounds such as legal, risk, compliance, data science, software development, design, user experience and ethics. Industry stakeholders should come together to create a road map for the development of such technologies and a set of principles to govern their use. Development of social care holograms like Lucy or social care robots will draw on expertise from various disciplines and will raise various ethical, legal, social

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

and economic issues. Hence, an ethics committee with multi-disciplinary expertise might well be the most appropriate to discuss the various issues raised.

- Policymakers should initiate public consultations about regulatory options governing mimicking technologies, especially where they are used to perform social care functions.
- Existing regulators should adopt a co-ordinated (co-regulatory) approach to AI mimicry to ensure harmonised, consistent rules for industry. As holograms like Lucy raise various issues beyond the remit of a single regulator, some mechanism is needed to ensure regulatory harmonisation.
- Governments and industry should encourage artists, directors, film producers, to create TV films or films showing positive and negative side of technologies that mimic people.
- No public funding should be directed to AI technology and human beings without the inclusion of ethics and SSH experts from the start. The EU should ensure that the ethical implications are included in every phase of the development, and public funding should therefore be directed at collaborative and multidisciplinary research on AI.
- We should consider how to find an optimal mixed of effective self- and co-regulation (which may also be embedded in the technique) and legislation as well as public supervision. To date, these are often considered as different topics, but the transnational scope of this topic requires a coherent system and integration of all these regulatory measures.

Information warfare

- National cybersecurity agencies should regularly inform the public about the scale and dangers of cyber warfare, attack vectors and possible responses. Other countries in the EU should emulate the actions of Estonia and Sweden to create “whole-of-nation” efforts intended to inoculate their societies from viral misinformation, including citizen education programmes, public tracking and notices of foreign disinformation campaigns and enhanced transparency of political campaign activities, so that citizens are informed about efforts to undermine their democracies.
- Countries should continue going public, identifying culprits, adopting sanctions. Public shaming of countries, “outing” their aggressive cyber warfare behaviour and activities should continue. Governments should reveal the full extent of cyberattacks, especially by states that operate without regard to international law or established norms and to do so with a feeling of impunity and without consequences. The US, UK and others should make strenuous efforts to cut off foreign powers’ ability to spread propaganda.
- It is not sufficient to merely expose a rogue state’s conduct, law enforcement authorities should seek to arrest those who break the law, but some malefactors may beyond the reach of the law of countries where cyberattacks occurred. Some retaliatory action is needed. For example, in the scenario, in retaliation to the shut-down of the two nuclear power plants in the UK in 2025, the US and UK could demonstrate their ability to turn off the power in the foreign power’s capital city with a one-minute black-out. They could threaten a longer black-out if the foreign power continues to attack their nuclear energy plants.[2] But other forms of retaliation are possible too, e.g., exposing the wealth of the foreign power’s leader hidden in the vaults of Zurich, the Cayman Islands and other such havens.
- The EC should provide funding for studies on information warfare via the European Defence Fund and the forthcoming Horizon Europe research programme and, in particular, how AI is being used to spread misinformation, hate crimes and lies, especially to undermine elections, and how to assess the resulting social impacts and what the EU should do about such activity.
- Compared with traditional armed conflict, the rules of information warfare are not well-defined. The European Commission and/or the United Nations should develop such rules, especially applicable to

[2] Ardehali, Rod, “Britain 'rehearses cyber-strike to black out Moscow' in the event of Russia attacking the West as thousands of UK troops stage biggest war-games exercise in a decade”, *The Daily Mail*, 7 October 2018. https://www.dailymail.co.uk/news/article-6248427/Britain-rehearses-cyber-strike-black-Moscow.html

the private sector. We need the information warfare equivalent of the Budapest Cybercrime Convention[3].

- The EC should promote a code of behaviour on the Internet. It should conduct studies indicating the real and opportunity costs wasted in cyberattacks and countering them.
- Politicians and diplomats should call for an end to information warfare, so that more resources can be channelled to combatting the effects of climate change.

Predictive policing

- To boost their trust with the public, policymakers should adopt a regulation making algorithms explainable to the public. Each algorithm should include a bit of code saying who created the algorithm, who paid for it, its purpose, website and contact for more information.
- Law enforcement authorities should ensure that criteria are clear and transparent for personal data to be entered into law enforcement databases.
- Policymakers should ensure there are independent regulatory authorities of sufficient size and clout to monitor the data in and use of law enforcement databases and offer commendations or impose penalties where appropriate.
- Decision-makers should ensure that measures in preventive policing and community investment supplement developments in predictive policing.
- Law enforcement authorities should take a balanced approach to local, white-collar and online hate crimes.
- Law enforcement authorities should ensure effective training of police officers and database operators in regard to the limitations of data analysis, particularly concerning the rates of false positives.
- The EU should sponsor research on automatically detecting when an attack is being planned and discussed

Self-driving vehicles: navigating towards an ethical future

- Governments should implement appropriate legislation and regulation on the sale, use and safety of SDVs, while national, international and supranational institutions should be responsible for ensuring that citizens are protected from the over-eagerness of manufacturers to put their vehicles on the road. The SDV automotive industry needs to be well regulated and controlled to ensure the safety of the vehicles through the effective implementation of SDV regulatory institutions.
- Clear delineations need to be established about what constitutes *essential data* for the vehicle's mobility and if this contains personal and private information. There needs to be clear indication that if essential data contains personal or private information, then it should be strongly anonymised, aggregated and secured to protect individual's privacy.
- Automobile manufacturers have the responsibility of identifying the purposes for which the car collects data in order to demonstrate their compliance with data protection law. For instance, there needs to be careful analysis if this data will be used for advertising, customised pricing or to sell additional products to the car owner, and either ensuring the owner is aware of these, and consent to it, or prohibit use of data in this way, altogether.
- Citizens should be informed about SDV regulation, so it is vital that policymakers receive input and feedback from the public about their needs. Policymakers should consider the needs of all stakeholders, so that policy is created for the public, rather than forced upon them by governments or SDV manufacturers.

[3] https://www.coe.int/en/web/cybercrime/the-budapest-convention

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

- For the foreseeable future, SDVs will have to use our current road signs, lights and markings to navigate on roads. However, these may eventually be replaced by ‘digital infrastructure’. Policymakers should ensure that there is a smooth transition between traditional infrastructure and the digital infrastructure of the future.
- SDVs offer great benefits for society, but also need to be carefully assessed and regulated before being integrated and used on our roads.

AI-powered education in 2025

- When designing AI-powered systems for education, conclusions or final decisions should not be made by the systems, even though the systems that support AI can also make intermediary decisions.
- Technologists and regulators should ensure human control over the use of AI, in order to eliminate many of the problems associated with fully autonomous systems. Such a requirement would protect the dignity of human life, freedom of choice, facilitate compliance with international humanitarian and human rights laws, and would promote accountability for unlawful acts.
- Considering that human behaviour does not always reflect human values, then AI systems, even though they are able to learn a lot by observing students and teachers, may be fundamentally unable to distinguish between value-aligned and misaligned human behaviour to provide AI educational products with appropriate learning feedback. A recommendation towards addressing such inconsistencies is to use a value-alignment mechanism to help systems distinguish between value-aligned and misaligned human behaviour.
- Educationalists must consider the design and implementation of new educational environments. We recommend the transformation of the current one-to-many teaching model of the classroom into facilitation environments, which focus on students achieving their learning goals by using project-based learning.

None of the scenarios discussed regulatory models or went into any depth on the nature of appropriate regulatory models, but all reflect the need for some form of regulation. The diversity of issues and applications illustrated by the scenarios suggest that regulation needs to be multidisciplinary in scope. One of the recommendations in the first scenario stated: “Existing regulators should adopt a co-ordinated (co-regulatory) approach to AI mimicry to ensure harmonised, consistent rules for industry. As holograms like Lucy raise various issues beyond the remit of a single regulator, some mechanism is needed to ensure regulatory harmonisation.”

Most regulators are sector specific[4], but AI crosses all sectors. To be effective, a regulator needs enforcement powers. A new regulator with a remit to challenge AI practices in whatever domain may lead to conflict with sector-specific regulators. So, when policymakers and legislators are thinking about regulatory options, they will need to take into account the sensitivities and the mandates of other regulators (where they exist).

Regulatory options are the subject of future SHERPA deliverables, but suffice it to say here, based on the scenarios and as an input to those later deliverables, that any new regulator or regulatory scheme will need to consider the inclusion of a wide range of competencies – technical, legal, ethical, organisational, economic, political, cultural – with enforcement powers across sectors and jurisdictions and with the sensitivities and diplomatic skills required to interact with other regulators, some of whom will already have formidable powers of their own.[5]

[4] The US Federal Trade Commission is an example of a regulator with powers that extend across many sectors in the economy.

[5] Interestingly, a few days after we wrote this comment, the House of Lords called for a super-regulator. See Hern, Alex, “House of Lords report calls for digital super-regulator”, *The Guardian*, 9 Mar 2019: “The House of Lords has called for

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Why our scenarios are an innovation, fit for purpose

Our scenarios and the methodology we used to create them offer value to policymakers who wish to engage stakeholders in a structured process considering future developments and their ethical, data protection, social and economic impacts.

To our knowledge, our structured approach to the scenario construction process is an innovation, yet it flows logically from the development of new technologies and applications to an illustrative vignette to the drivers, the inhibitors, and the ethical, data protection, social and economic impacts. The scenarios conclude with some recommended measures to reach a desired future and avoid an undesired future.

Our scenario construction methodology is based on engaging with stakeholders from the get-go, from an initial brainstorming workshop through several iterations of the scenario. Part of the reason to invite increasingly greater numbers of stakeholders to review and comment on the scenario is to prompt stakeholders to consider the implications of advanced new AI technologies, the risks and benefits. In other words, construction of a scenario is also an awareness-raising exercise. However, ultimately, a scenario is a policy-making tool, and our scenarios are constructed in a way so that policymakers can readily grasp their import and can use the scenario methodology themselves on other issues.

Partners and stakeholders who put the scenarios together

All SHERPA partners were involved in the development of these scenarios, some more so than others. The scenario construction processes were led by TRI (first and second scenario), UT (third and fourth scenario) and UCLANCY (fifth scenario).

We acknowledge and thank our stakeholder board members and stakeholders from the project's contact list for their participation in the workshops and their comments and suggestions during and after those workshops.

Next steps

This deliverable is intended to be a "living" document. The scenarios herein have already benefitted from a wide range of stakeholders who were invited to the scenario brainstorming workshops or are members of the SHERPA stakeholder advisory board or selected stakeholder experts. Although we are now submitting the deliverable to the European Commission, we are also inviting comments from our contact list and from visitors to the SHERPA website. We welcome comments from as many stakeholders as possible and will take their comments into account in subsequent revisions of the scenarios until August 2019, after which we will consider the scenarios to be definitive.

The individual scenarios and the executive summary of this deliverable will be integrated into the SHERPA Workbook which will be available for consultation on the project website.

In the workbook, we offer three levels of detail and stakeholders are welcome to pick whatever level they prefer. The first level is an infographic that provides a thumbnail sketch of each scenario. For those who want

the creation of a digital super-regulator to oversee the different bodies charged with safeguarding the internet and replace the "clearly failing" system of self-regulation by big technology companies. A new Digital Authority is the chief recommendation of the Lords' communications committee report, which warns that the patchwork quilt of more than a dozen regulators that oversee the digital realm creates gaps and overlaps." The chair of the committee, Lord Gilbert of Panteg, said, "Self-regulation by online platforms is clearly failing and the current regulatory framework is out of date. The evidence we heard made a compelling and urgent case for a new approach to regulation. Without intervention, the largest tech companies are likely to gain ever more control of technologies which extract personal data and make decisions affecting people's lives."

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

more detail, they can click on a link that takes them to the executive summary and for those who want even more detail, they can click on a link that takes them to the full scenario. With each scenario, we include several questions inviting the views of visitors to our website.

To facilitate comments on the scenarios, a little robot icon asking questions will appear as the reader scrolls down the page of each scenario. It will be clickable and have a speech bubble prompting visitors to provide their feedback.

We will provide the Commission with two revisions of this report taking into account the comments received. We may not reflect all comments in the scenarios – some may not be germane – but we will certainly review and consider all comments that can enhance the scenarios and our recommendations to policymakers. The first revision will be in June and the last revision will be in August 2019. While we will still welcome comments on the scenarios after August, we will not make any further revisions beyond that date.

List of figures

Figure 1: Each step in the scenario construction process feeds into the next

Figure 2: First scenario

Figure 3: AI assistants - Image Credits: Tony Webster, Flickr (CC BY 2.0)

Figure 4: Data breaches

Figure 5: Second scenario

Figure 6: Phishing

Figure 7: Cyber attacks

Figure 8: Cyber weapons

Figure 9: Cybersecurity

Figure 10: Third scenario

Figure 11: CCTV - Image Credits: Niv Singer, Flickr (CC BY-SA 2.0)

Figure 12 Police investigations

Figure 13: Fourth scenario - Image credits: smoothgroover22, Flickr (CC BY-SA 2.4)

Figure 14 The joy of driving

Figure 15: Fifth scenario

Figure 16: Classroom - Image Credits: Gokhan Okur freeimages

Figure 17: Image Credits: Kingwood freeimages

Figure 18: istock, bought by DS

Figure 19: Image Credits: Jean Scheijen freeimages

Figure 20: Concerns for Governments

Figure 21: Ethical challenges

List of tables

Table 1: List of acronyms/abbreviations

Table 2: Glossary of terms

List of acronyms/abbreviations

Abbreviation	Explanation
SIS	Smart Information Systems
AI	Artificial Intelligence
ICT	Information and Communications Technology
SDV	Self-driving vehicles
SSH	Social Sciences and Humanities
CCANP	Citizens Committee Against Nuclear Power
HE	Horizon Europe
GDPR	General Data Protection Regulation
ERAAI	European Regulatory Agency for AI
LEAs	Law enforcement authorities
ePR	ePrivacy Regulations
VR	Virtual Reality
AR	Augmented Reality
IoT	Internet of Things

Table 1 List of acronyms/abbreviations

Glossary of terms

Term	Explanation
Stakeholder	A relevant actor (persons, groups or organisations) who: (1) might be affected by the project; (2) have the potential to implement the project's results and findings; (3) have a stated interest in the project fields; and,

Term	Explanation
	(4) have the knowledge and expertise to propose strategies and solutions in the fields of SIS and artificial intelligence (AI)
Scenario	A tool for ordering one's perceptions about alternative future environments in which one's decisions might be played out concretely, so people can help people make better decisions[6]
Delphi study	Expert survey in two or more 'rounds' in which, in the second and later rounds of the survey the results of the previous round are given as feedback.[7]
Backcasting	a planning method that starts with defining a desirable future and then works backwards to identify policies and programs that will connect that specified future to the present[8]
Deepfake	Deepfake is an AI-based technology used to produce or alter video content so that it presents something that didn't, in fact, occur[9].

Table 2 Glossary of terms

[6] Wright, D. et. al, 2013

[7] Cuhls, K., "The Delphi method", Undated. https://pdfs.semanticscholar.org/21a4/a0ac70928452880eae6c51e171aa9289a00a.pdf

[8] https://en.wikipedia.org/wiki/Backcasting

[9] https://whatis.techtarget.com/definition/deepfake

1. Objectives of this deliverable

Our objectives for this deliverable are set out in the Description of Action attached to the SHERPA Grant Agreement with the European Commission. It states that “In Task 1.2, SHERPA will develop five scenarios exploring emerging SIS… highlighting critical ethical and human rights issues arising from the use of future SIS. The scenarios will also highlight cyber-security risks”. The task description adds that “The partners will engage stakeholders in the construction and validation of the scenarios. The partners will disseminate the scenarios and the scenario methodology to a range of stakeholders asking for their views on minimising the risks to ethics and human rights while maximising the economic and societal benefits of the new technologies.”

With these objectives in mind, we set out our results in the following five pages. Following the introduction, we present five scenarios and finish this report with our conclusions.

2. Introduction

In this report, we posit a new methodology for constructing scenarios specifically for policymaking. Policy scenarios provide a useful methodology to engage an increasing number of stakeholders in exploring the issues expected to influence the development and take-up of emerging technologies, such as artificial intelligence, and to arrive at a consensus about a socially acceptable future regarding the technology and desirable and undesirable pathways for achieving it. However, many scenarios and scenario methodologies do not follow a structured approach. The structure of the scenarios created using the methodology put forward here highlight ethical, legal, social and economic issues that are relevant for policymakers. Using this approach, policymakers can create various scenarios that enable comparative analyses. For most other types of scenarios, policymakers need to deconstruct the scenario in order to determine what is relevant for policymaking. For the policy scenarios described here, it is not necessary to deconstruct them because they are structured in such a way as to facilitate policy analysis. Also of value to policymaking, and unlike many other scenario methodologies, our approach is specifically designed to engage stakeholders from the beginning to the end of the scenario construction process. Indeed, the legitimacy of our scenarios stems from inviting increasing numbers of stakeholders to comment on each iteration of the scenario. These and other features make the policy scenario methodology both innovative and well-suited to be an instrument of policymaking.

The chapter outlines how to construct policy scenarios and how to generate comments from stakeholders, especially to help identify and discuss the ethical, human rights, data protection and social, economic and other consequences arising from the application of emerging technologies. This chapter is followed by five scenarios concerning artificial intelligence in different applications in 2025. We used our methodology to construct the scenarios.

We recommend that policy scenarios be placed in a timeframe five to seven years away (2025[10] in the instance of our scenarios). This will generally give policymakers enough time to act on the recommendations

[10] In setting 2025 as the time frame of our scenarios, we subscribe to the advice offered by Cairns & Wright that “an appropriate future timeframe should not extend so far into the future as to require ‘science fiction’ thinking or be so close that the future is fairly predictable. It should represent a reasonable long-term planning horizon in relation to the

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

that emerge from the scenario, while providing a not-too-distant horizon for engaging non-experts. We believe that it is both possible and necessary to create plausible scenarios. Plausibility is important for policymakers. If the scenarios were positioned much later – say 2035 – policymakers would not feel the same urgency to respond to the recommendations. Near-term plausibility is more likely to pressure policymakers to act than a situation a decade or more away. Plausibility as a criterion is also important for engaging non-experts. Plausibility emphasises the relevance of the scenario and invites people to think how such scenarios might relate to their own lives in the future.

With policy scenarios, we seek to achieve the greatest possible consensus. Plausibility and probability are both important to achieving consensus. If stakeholders view the scenarios as plausible and probable, they are more likely to lend their weight to the recommendations that emerge from the scenarios. However, having said that, one should not discard outlier opinions. Rather, they can be noted in the scenario, especially where contrary views offer some useful insights and depend on contextual factors that present critical uncertainties.

Why use scenarios?

Scenarios are used for a variety of purposes. The so-called father of modern scenario construction, Herman Kahn, used scenarios for strategic planning and war games while he was with the RAND Corp in the 1950s. In the following decade, as director of the Hudson Institute, a non-profit research institute, he used scenarios for issues related to US public policy, international development, and defence.[11] Shell became an early adopter of scenarios for predicting the volatility of oil prices and the future of the Soviet Union, a big competitor in the oil market. The UN International Panel on Climate Change (IPCC) has used scenarios for many years to predict the impacts of climate change.

The use of scenarios in policymaking is well established.[12] The use of scenarios as a participatory exercise is also well established.[13] Others have pointed to still further uses: e.g., the use of scenarios as techniques that can bring together different types of knowledge and act as platforms for knowledge brokerage.[14] Just as scenarios serve different purposes, so there are different types of scenarios, among which are the following.

Types of scenarios

Some of the principal types of scenarios are:

i. **Best case, status quo, worst case** – The scenario authors create three scenarios: one describes a best-case scenario, the second describes the status quo (i.e., a future is based on a continuation of present trends), and the third describes a worst case. This was the approach that Herman Kahn

organizational response." See Cairns, George, and George Wright, *Scenario Thinking*, second edition, Palgrave Macmillan, 2018, pp. 34-35. On p. 40, they mention a timeframe of seven to 10 years.

[11] Glenn, Jerome C., and The Futures Group International, "Scenarios" in The Millennium Project, Futures Research Methodology—V3.0 [n.d.], p. 1.

[12] Here is an example: Volkery, Axel, and Teresa Ribeiro, "Scenario planning in public policy: Understanding use, impacts and the role of institutional context factors", *Technological Forecasting and Social Change*, Vol. 76, Issue 9, Nov 2009, pp. 1198-1207. Here is another example: Groves, David G., and Robert J. Lempert, "A new analytic method for finding policy-relevant scenarios", *Global Environmental Change*, Vol 17, Issue 1, February 2007, pp. 73-85.

[13] And here is yet another example, among others: Duckett, Dominic George, Annie J. McKee, Lee-Ann Sutherland, Carol Kyle, Lisa A. Boden, Harriet Auty, Paul R. Bessell, Iain J. McKendrick, "Scenario planning as communicative action: Lessons from participatory exercises conducted for the Scottish livestock industry", *Technological Forecasting and Social Change*, Vol. 114, Jan 2017, pp. 138-151.

[14] Collingwood Environmental Planning (CEP) and Milieu Ltd, BLOSSOM — Bridging long-term scenario and strategy analysis: organisation and methods: A cross-country analysis, EEA Technical report No 5/2011, European Environmental Agency, Copenhagen, 2011.

followed.[15] A strength of this approach is that each scenario can take into account different factors, or the intensity of a range of factors. Such scenarios are not optimal for policymakers, as they offer three different outcomes depending on what steps stakeholders do or do not take to reach a desired future or avoid an undesired one. By contrast, our policy scenario methodology aims to achieve a consensus on what steps to take.

ii. **Orthogonal futures** – The approach is based on a matrix containing four quadrants. Each quadrant represents a permutation of the future based on two key driving factors represented by the X and Y axes. The X and Y axes can also represent likelihood (low to high) and impact (low to high). A weakness of the orthogonal approach is that it is driven by two axes, two factors. Hence, orthogonal futures may lead to somewhat formulaic, over-simplistic views of the world, skimming through more nuanced factors and complex interactions inherent in stakeholder insights.[16]

iii. **Dark scenarios**[17] – This approach focuses on a future where things go wrong. However, with a policy scenario, we recognise that an emerging technology will have some good points and bad points, some pluses and minuses. Hence, dark scenarios serve as warnings – at best, they steer us away from negative futures, but give little guidance on what steps to take to reach a desired future.

iv. **Ethical dilemma scenarios**[18] – This approach focuses on a future where there is no obvious correct course of action. The scenario sets out a dilemma and provides a basis for discussion by stakeholders to reach a consensus on a course of action. By contrast, a policy scenario already represents a consensus of stakeholders.

v. **Narrative scenarios** – tell a story. These stories are often in the form of scripts that include roles and can be very elaborate, as in the film *Minority Report*.[19] Narrative scenarios can be confining in the sense that they focus on telling a story and may not allow the authors (the stakeholders) to explore all ethical, legal, social and economic aspects raised by the new technologies. Moreover, for the sake of narrative consistency, this type of scenario may gloss over existing contradictions.

vi. **Policy scenarios** – This approach describes a future and results in recommendations that policymakers and/or other stakeholders should implement, and the steps that stakeholders should take, to reach a desired future and avoid an undesired future. This approach seeks consensus with stakeholders about how to manage the issues raised in the scenario. Policy scenarios are designed

[15] Glenn, op. cit.

[16] Ramirez, Rafael, Malobi Mukherjee, Simona Vezzoli and Arnoldo Matus Kramer, "Scenarios as a scholarly methodology to produce 'interesting research'", *Futures*, Vol. 71, 2015, pp. 70–87.

[17] Wright, et al., coined the term "dark scenario" in the context of the four dark scenarios created as part of the EU-funded SWAMI project. The scenarios and the methodology are described in Wright, David, et al., *Safeguards in a World of Ambient Intelligence*, Springer, Dordrecht, 2008.

[18] Wright, et al., September 2014.

[19] Other examples of narrative scenarios are those exploring the benefits of ambient intelligence. "Undoubtedly the best-known AmI scenarios are those produced for ISTAG, a group with about 30 members from industry and academia, which advised the European Commission's Information Society Directorate General. In May 2000, ISTAG commissioned the creation of four scenarios 'to provide food for thought about longer-term developments in Information and Communication Technologies', with the intent of exploring the social and technical implications of AmI. The ISTAG scenarios were actually developed by the Institute for Prospective Technological Studies (IPTS), which is part of the European Commission's Joint Research Centre, in collaboration with about 35 experts from across Europe." Wright, 2008, op. cit.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

to explore the key factors (drivers) affecting ethical, legal, social and economic aspects of everyday life in the future (2025 in the case of our scenarios). They provide a shared view on the issues arising from the emergence of particular technologies and on recommendations to address those issues, so that we reach a desired future and avoid an undesired future. The technologies are grounded in plausibility so that we can tease out the ethical and social issues that we think will confront policymakers and other stakeholders in the near future. Our scenarios are examples of policy scenarios.

Scenarios can also be categorised or described in terms other than those above. Most of the scenarios described above could also be of an exploratory or normative nature, or an instance of backcasting. They could also be of the following types.

Trend scenarios (sometimes called reference, extrapolative or predictive scenarios) start from the present and project forward on the basis of to-be-expected trends and events [forecasting]. They are intended to be realistic rather than, for instance, normative or extreme.

Normative scenarios are developed to evaluate how a specific outcome can be reached. They are designed on the basis of a set of desired features or 'norms' that the future world should possess. The exercise then consists of tracing backwards [*backcasting*[20]] a viable path from such an outcome to today—pointing the way to reach that desired future. Normative scenarios often reflect more radical discontinuities; they can be combinations of technological possibilities and political ambitions or targets.

Exploratory scenarios are explorations of what might happen in the future. They are based on identifying critical uncertainty factors and on different expectations of technical and/or policy developments over the near- to medium-term. [21]

Our policy scenario approach is different from others in that we offer a structured approach, from brainstorming on how technology and applications might evolve over the medium term (six or seven years) to formulating recommendations to policy-makers. Also, increasing the number of stakeholders invited to comment on the scenario is a key element in our approach. Another difference is that we do not focus on developing a set of scenarios (see below for a brief description of the different types), rather we aim to achieve the greatest possible consensus on a single policy scenario.

Policy scenarios serve policymakers as policy input and stakeholders as a participatory exercise. Our type of scenario supports policymakers in several ways:

- To explore possible consequences of current trends
- To engage stakeholders
- To uncover issues that might otherwise be overlooked
- To help decision-making
- To consider desired and undesired futures
- To determine what steps should be taken to reach the desired future and avoid an undesired future.

Policy scenarios can complement other policymaking instruments such as Delphi studies, which engage a small group of selected experts. Recommendations emerging from scenarios constructed by engaging a wide

[20] Vergragt, Philip J., and Jaco Quist, "Backcasting for sustainability", *Technological Forecasting & Social Change*, Vol. 78, 2011, pp. 747–755.

[21] Wright, Sept 2008, p. 474.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

range of stakeholders, including the public, can be compared with those emerging from smaller groups of experts.

Why consult stakeholders?

Consultation is "necessary to engage more widely with the public as well as with stakeholders, to increase the accountability of the process by providing better opportunities for the public to participate early and effectively in major policy decisions (in line with the Aarhus Convention)".[22] Engaging stakeholders is a fundamental requirement in the construction of policy scenarios of the type outlined in this article. It is important to engage diverse stakeholders in the scenario construction process from the outset to ensure that the forecasting is representative of a variety of perspectives. There are several reasons why this should be so:

- Stakeholder buy-in is achieved by involving stakeholders right from the beginning of the process of constructing the scenarios.
- Stakeholder involvement is important for credibility of the scenarios. If stakeholders are involved in the scenario construction process from the outset, the prospect of credible scenarios is much higher than if they are not involved. By inviting increasingly large numbers of stakeholders to review and comment on the scenarios, we increase the credibility and legitimacy of the scenarios.
- By engaging stakeholders, we gather ideas that might not otherwise have occurred to us. Scenario construction should be a multi-disciplinary exercise.
- Stakeholders can help to disseminate the scenarios.

A critical success factor for gaining consensus around the scenario is an adequate contact list, preferably with hundreds of diverse stakeholders who can be invited to comment on the scenario(s).

Policy scenarios are most suited for reaching a consensus on the ethical, legal, social and economic drivers, impacts, barriers, issues and recommendations to reach the desired future and avoid the undesired future. However, the same scenario methodology could be used for exploring social issues, e.g., racial discrimination or budgetary squeezes on the national health service or deforestation.

Kicking off with a scenario workshop

The first step in constructing a policy scenario is to identify stakeholders to participate in an initial brainstorming workshop.[23] The workshop provides key information and ideas on which the scenario is drafted and opened up to consultation. Hence, it is desirable to have workshop participants from diverse backgrounds and disciplines, who will help ensure that different viewpoints are taken into account in constructing the scenario. An optimum number is 10 – 15 participants. We want all workshop participants to be actively engaged in brainstorming on the development of the scenario, which becomes more problematic with greater numbers.

How to do it…

In advance of the workshop, participants should be encouraged to do their own literature search around the topic and suggest any background reading. They should also be sent some articles about the technologies on which the scenarios focus. Giving workshop participants some homework in advance of the workshop is helpful to make sure that participants are more or less starting from the same vantage point and a common vocabulary with which to communicate their positions. To avoid the risk that one article might inject a bias

[22] Ibid., p. 62.

[23] The policy scenario development methodology draws on the basic scenario development approach advocated by Cairns & Wright, op. cit. It also draws on the approach developed by Wright, et. al, op. cit., 2008. Its distant antecedent is the approach advocated by Peter Schwartz in his classic *The Art of the Long View* (Doubleday, 1991).

in the scenario development, the workshop leader should provide two or three or more articles from different, credible sources.

As most or all of the participants may not know one another, brief introductions should be made at the outset. It is helpful for participants to understand what type of scenario is to be constructed and why. As our policy scenario is innovative in structure (while building on the results of others) and engagement with increasing numbers of stakeholders, the scenario leader needs to explain why we have chosen a policy scenario to explore the implications of smart information system technologies in 2025.

Workshop participants will benefit from being explicitly informed about what is meant by a plausible scenario and how they are supposed to envisage such a future. At the outset of the scenario workshop, the workshop leader should provide an introduction to the purpose of brainstorming, and what participants should aim to achieve during the day. It is important to emphasise that Chatham House rules will apply, i.e., views expressed during the workshop will not be attributed to particular participants.[24] In the European Union, it is important that workshop participants be given an information sheet and informed consent form to sign before the workshop or at the outset of the workshop.[25] The information sheet should make it clear that stakeholder views will not be ascribed to specific participants.

Following introductions and a presentation of the workshop's objectives and its agenda, the workshop leader might wish to invite a technology expert to give an overview of the state of the development of the technology and to speculate where the technology might be in several years from now. For example, a senior official from F-Secure gave a presentation on AI and technologies that mimic people that kicked off the discussion at our first workshop.

The initial workshop should cover key questions for scenario workshop participants, the first of which is:

- **Expected technological progress:** Where will the technology be by the time horizon specified (e.g., 2025 in our case)? How advanced will it be? In what applications will it be used? Workshop participants should be encouraged to stretch their imaginations about how technology will develop and be used six years from now. Participants should consider how fast technology has developed in recent years, considering the advances in the past several years in smart phones, social networks, the Internet of Things, e-commerce, wearable technology, facial recognition, robots, augmented reality, virtual reality, 3D printing, drones, etc. and how fast it might develop in the next six years. It is important to allow sufficient time for this first brainstorming session, as it will provide the context for the brainstorming that follows. The workshop leader must steer the discussion towards the advance of technology and its applications rather than let it drift toward topics that are on the agenda for later, e.g., ethical issues. Through the brainstorming on the technology and its application, workshop participants can reach a consensus on the application(s) that will feature in the vignette, which aims to illustrate how the technology might be used.

Through the discussion of where the technologies and their applications might be in the future, the workshop participants should discuss a vignette that illustrates the use of such technologies several years hence. If participants give appropriate consent, it will be useful to record the brainstorming discussion as a reference for constructing the scenario. As a minimum, the workshop leader should have someone who can take detailed notes of the discussion.

The next key questions in the scenario construction process should be:

[24] Chatham House rules signify that participants can say whatever they want, knowing that they will not be quoted by name. https://www.chathamhouse.org/chatham-house-rule

[25] This is an ethics requirement, independent of methodology and relationship between participants.

- **Drivers:** the second step in the scenario construction process is to consider the drivers that impel the development of these technologies and their applications. Drivers can be social, economic (or financial), political, regulatory, technological, etc. Workshop participants should discuss the relevance of each of the different drivers identified.
- **Barriers and inhibitors**: Third, workshop participants should consider the potential barriers and inhibitors for the production and uptake of these technologies in the time horizon specified. While some barriers are foreseeable, others appear unexpectedly. So, workshop participants should consider whether there are externalities or "black swan" events that may impact the development of the technology. A black swan event might be Silicon Valley and the rest of California descending into the Pacific Ocean as a result of a major earthquake along the San Andreas fault.

There are two different approaches for identifying drivers. The first is the traditional approach, as advocated by Cairns & Wright[26]. In this approach, each of the workshop participants identifies five drivers on sticky notes, which can be affixed to a wall. The drivers can be just a few words or a sentence but should be self-explanatory and clear as to what it means, e.g. 'saves time'. We suggest the workshop leader ask participants to each suggest five drivers, which is a manageable number for most participants. A larger number would leave participants contributing less important or inconsequential drivers simply to fill their quota. The first to finish writing his or her drivers can place them horizontally on a wall. As the other participants finish writing their drivers, they can see what has already been fixed and add their notes in column formation to any that are similar or, if not, create a new column. The various sticky notes can then be grouped or clustered under some key words that describe those under each column. We are especially interested in those regarded as the most important cluster(s) of drivers in terms of their impact and those that reveal the greatest certainty.

A second approach to scenario construction is to divide participants into groups of three or four and to have two of these groups exchange views and identify drivers, and two other groups identify barriers or inhibitors to the development and application of the technologies under discussion. Each of the groups then reports to the plenary, where there is further discussion of the drivers and barriers until participants reach a consensus on the principal drivers and barriers.

The policy scenario should reflect consensus on the drivers that will impel development of the technology and its applications, as well as the barriers that might inhibit such development. We identified several key drivers and barriers in our mimicking scenario.

The final steps to scenario construction in the initial workshop are:

- **Impacts:** The participants should consider the ethical, legal, social and economic impacts of these technologies and their applications (in 2025). Workshop discussion should be structured according to each of these impacts.
- **Recommendations**: Workshop participants should debate how we as a society will be able to mitigate the negative and accentuate the positive impacts of these technologies in 2025. What steps can we and/or policymakers take to reach the desired future and avoid the undesired future? What recommendations should we make to policymakers or other stakeholders?
- **Next steps**: What steps should be taken to engage stakeholders and achieve buy-in of the scenario created and its recommendations?

[26] Op. cit.

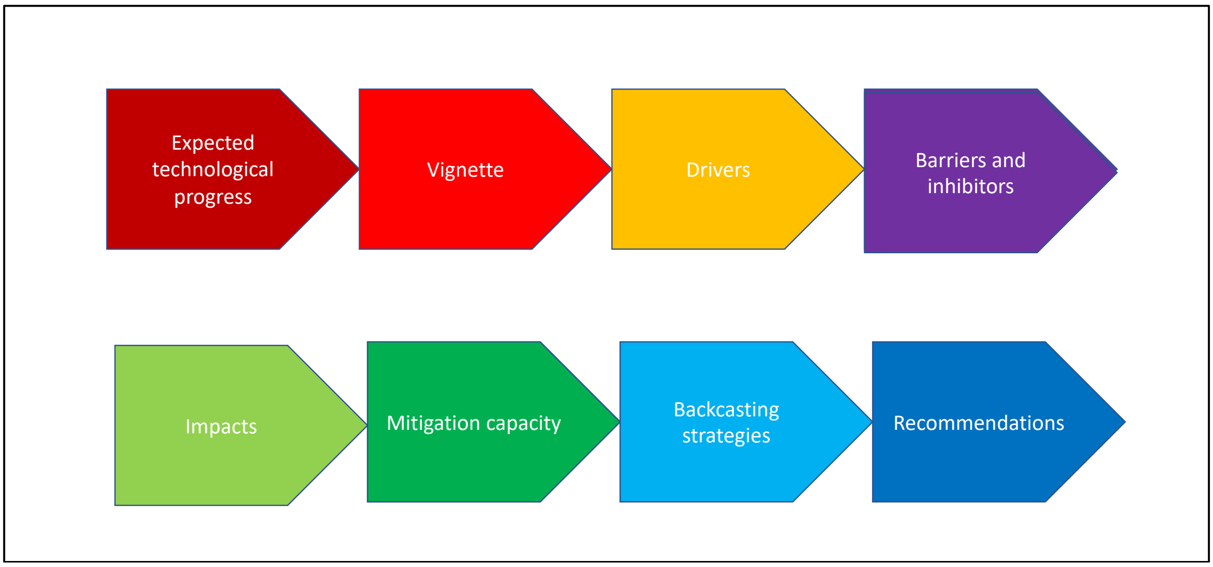


Figure 1 Each step in the scenario construction process feeds into the next

Constructing the scenario

A policy scenario should be structured to address these questions. In our case, the scenarios were typically structured with these headings;

- Introduction – purpose of the scenario, how it was constructed
- A description of the technology (or cluster of technologies) in 2025
- Typical applications in 2025
- A brief vignette to illustrate the topic on which the scenario focuses
- Drivers of the technology
- Potential barriers to and inhibitors of the technologies in 2025
- Ethical, legal, social and economic impacts in 2025
- Mitigating the negative, accentuating the positive – recommendations to policymakers and other stakeholders to reach the desired future and to avoid the undesired future.

With policy scenarios, we do not aim to predict a specific future. That is impossible. However, we can envisage a plausible future (the vignette) and the many factors – the drivers, barriers, impacts – that policymakers should take into account to enable or avoid a future like that envisaged in the scenario. We do not want a scenario with many variables and possible turns of event. Policymakers prefer a clear, single course of action that has stakeholder support. In addition to the structured approach to its development, our scenario shows the range of factors that policymakers should also take into account in the formulation of policy and recommendations.

Reaching out to stakeholders with new iterations

Constructing scenarios is an iterative process. It is analogous to throwing a stone in still water. As the stone strikes water, it generates a series of ripples radiating outwards with each ripple (or circle) bigger in size than its predecessor. In the same way, we want to engage an increasing number of stakeholders with each iteration of the scenario.

Therefore, with each iteration of a policy scenario, we disseminate it to a wider group of stakeholders, from experts to the public. We want to gather comments from stakeholders until the scenario is stable, i.e., we have resolved most if not all stakeholder comments and issues, and the remaining stakeholders have no or few minor comments. It would not be appropriate to disseminate the first draft of a scenario to a large number of stakeholders; sending a first draft that has no significant support from any stakeholders is an unfair imposition on a larger group of stakeholders. A scenario sent out to a large number of stakeholders for comment should already represent a consensus among a smaller number of stakeholders.

We suggest a maximum of four iterations with stakeholders.

1st iteration

The first iteration of the scenario emerges from the first stakeholder workshop, where we delineate the issues to be explored in the scenario. Taking into account the discussion during the workshop, the scenario leader drafts an initial scenario. The leader sends the first draft of the scenario to workshop participants to gather their comments and then prepares a second iteration. The draft scenario should desirably be fewer than 10 pages in length, so as to encourage policymakers and other stakeholders to review and comment on it.

2nd iteration

The scenario leader sends the second iteration of the scenario to a larger group of stakeholders, perhaps 30 or so, to seek their comments.[27] In addition to the second iteration, the scenario leader may wish to include some questions, mainly qualitative, to stimulate responses from stakeholders. The questions should relate to each principal section of the scenario. The scenario leader gathers comments from the stakeholders, which are used to prepare a third iteration of the scenario.

3rd iteration

The scenario leader sends the revised scenario, together with the questions, to a much larger contact list.[28] The leader can use various methods, including e-mailing the scenario to stakeholders on the contact list or refer them to a webpage where they can respond to questions.

A word of caution: at this point the scenario leader must be careful to avoid the scenario becoming a mish-mash of comments and amended text. The policy scenario must be coherent, must make sense, and must be easily readable.

4th iteration

Following receipt of comments from stakeholders on the contact list, the scenario leader revises the scenario again (the fourth iteration), posts it on the project website and invites comments from visitors (the wider general public) to the website. The scenario leader should fix a cut-off date for comments. The scenario leader then revises the scenario one last time, at which point the scenario can be regarded as stable, containing a set of policy, regulatory, technical, organisational and/or other recommendations to address the challenges raised by the technologies.

[27] In the EU-funded SHERPA project for which this policy scenario was prepared, we sent the scenarios to the project's stakeholder advisory board, comprising 28 stakeholders, inviting their comments.

[28] In the instance of the SHERPA project, we invited about 800 people on the project's contact list to offer their feedback on the scenarios.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Discussion

Policymakers will be interested in recommendations that are both credible and useful. And to be credible and useful, scenarios have to respect five conditions; *pertinence*, *coherency*, *likelihood*, *importance* and *transparency*.[29] Our scenario (Annex) meets these five conditions:

- The subject matter is *pertinent*. Technologies such as those presented in the scenarios are already receiving much media attention and raise many issues. The technologies are all relatively accessible and affordable in 2025, even the holograms mentioned in the first scenario or the SDVs mentioned in the fourth scenario and even the learning buddies mentioned in the fifth scenario.
- The scenario is *coherent*, as illustrated by the vignette. The scenario follows a logical sequence from the introduction to the technology and its applications, through the drivers, barriers, impacts and finishing with recommendations to policymakers.
- The scenario's *likelihood* can be debated, but as Durance and Godet point out[30], a scenario is not a future reality. This scenario is plausible, as all of the technologies that underpin the scenarios are in development in 2019.
- The scenario is *important*, as it raises numerous ethical, legal, social and economic matters, including risks to privacy and data protection.
- The scenario fosters *transparency*. It emerged from a workshop of stakeholders and increasing numbers of stakeholders have been invited to comment on the scenario. Such an open invitation is an indicator of transparency.

In addition to meeting the above conditions, the technology policy scenario builds on existing scenario methodologies. The structure of policy scenarios, as postulated here, resonates well with another point made by Durance and Godet. They note that "Morphological analysis… has become among the most popular tools. … it lends itself perfectly to the construction of scenarios. Using morphological analysis, a …system can be decomposed into dimensions … demographic, economic, technological, and social/organizational."[31] Our policy scenario approach is, in effect, a morphological analysis of the factors several years hence that affect the development and deployment of a new technology that meets a social need (e.g., monitoring the well-being of senior citizens, urging citizens to be careful not to get caught in the crossfire of information warfare, predicting crime before it happens, reducing pollution with self-driving cars and improving education). However, although the technologies meet social needs, they also pose threats. The structure and process that we use to create a technology policy scenario facilitate the identification of such issues and ways in which to address the risks. Our process is open and transparent. It seeks out comments from an increasing number of stakeholders. Our scenario construction process meets the final test set out in Durance and Godet's article, i.e., "the complexity of the problems and the need to address them collectively require methods that are as rigorous and participatory as possible so that the individuals involved may identify the appropriate problems and agree upon solutions."[32] Our scenario construction process is inclusive and participatory. It seeks out engagement with stakeholders from the initial brainstorming workshop with stakeholders, through to posting the scenario on the project website and inviting comments from the wider public.

Getting the scenario in front of policymakers

When the scenarios are stable, the scenario developers will want to get as many policymakers as possible to review them and consider (we hope) the recommendations. This is the way to create impact with scenarios.

[29] Durance, Philippe, and Michel Godet, "Scenario building: Uses and abuses", *Technological Forecasting & Social Change*, Vol. 77, 2010, pp. 1488-1492 [p. 1488].

[30] Ibid.

[31] Ibid., p. 1490.

[32] Ibid., 1491.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

However, getting policymakers to focus on a scenario and its recommendations is a serious challenge, both at the European Union level as well as at the national level.

Policymakers will be interested in recommendations emerging from a stakeholder group where there is a consensus. Of particular interest, of course, will be those instances where there is a congruence of recommendations or where the recommendations are similar.

How recommendations cohere across other foresight tools will be important. Policymakers will more likely be interested in recommendations where there is congruence with the different approaches (case studies, scenarios, Delphi studies, etc.).

There are different ways to draw policymakers' attention to a scenario. We suggest the following as an outline framework:

Direct contact with policymakers

- Involve policymakers in the process of commenting on the third iteration.
- Snowballing – ask an initial group of policymakers to forward the scenario to their contacts
- Compile a list of relevant policymakers in the European Commission and Member States and commend the scenario in any e-mail to them (preferably individually addressed). Emphasise wide stakeholder support for the scenario and its recommendations.
- Hold meetings with key policymakers and give them a face-to-face briefing.

Generate public and media interest

- Feature the scenarios on the project's website.
- Generate interest in the scenarios and recommendations by sending press releases to the media.
- Use Twitter to inform stakeholders about the scenario.

Academic dissemination

- Prepare one or more scenario-based articles for peer-reviewed journals.
- Etc.

We now turn to the five scenarios. As mentioned above, they have a similar (though not rigidly identical) structure and objectives. We want to explore the issues that might arise with the emergence and deployment of these new technology clusters, their impacts and recommendations to policymakers to help deal with those impacts, to avoid an undesired future and to reach a desired future.

3. First scenario: Creating companions for senior citizens with technologies that mimic people

Figure 2 First scenario

Creating companions for senior citizens with technologies that mimic people

1. Introduction

This scenario focuses on technologies and applications that mimic people, and that are used to create companions for senior citizens in the year 2025.[33] They work by feeding hundreds or thousands of images of a person's face or body into a machine-learning algorithm that then maps them onto video of another person's body. Anything the person in the video does or says can be made to look like it is coming from the target.[34] Similar algorithms can be used to replicate a person's voice, to make it seem as if the target person is saying something that in fact the

[33] But such technologies can be used in many settings -- in pornography, in movies, in political campaigns, in security, in the provision of public services and much else.

[34] Hawkins, Derek, with Bastien Inzaurralde, "Doctored videos could send fake news crisis into overdrive, lawmakers warn", *The Washington Post*, 31 July 2018. See also Chesney, Robert, and Danielle Keats Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security", *California Law Review*, Vol. 107, 2019 (forthcoming). A pre-publication version of the article can be found here: https://papers.ssrn.com/sol3/results.cfm

target never uttered. Previously, technology used a large database of recordings of one person's voice uttering a long collection of sentences, selected so that the largest number of phoneme combinations were present. Synthesising a sentence was done just by stringing together segments from this corpus.[35] More recently, artificial intelligence is making human speech as malleable and replicable as pixels. Lyrebird, a Canadian start-up, uses a set of algorithms that it claims can clone anyone's voice by listening to just a single minute of sample audio.[36]

Like the Internet of Things and augmented reality, artificial intelligence is blurring the boundaries between the digital and physical worlds. This scenario concerns a couple. Alfred is a real person, and Lucy is a hologram, the manifestation of several key technologies, including machine learning, big data analytics, artificial intelligence, facial recognition, audio recognition, IoT sensors and actuators, augmented reality, virtual reality and, not least, holograms. In fact, naming *a* technology that mimics people in the singular is problematic because there is not just one technology but many. Our scenario depicts a future in 2025 when we see a cluster of technologies working together – technologies that mimic the voice, the image, the behaviour, the gait and movement of a person, an avatar that knows the history of the target person.

Our scenario considers the ethical, legal, social and economic issues that arise from the use of such technologies and the steps we, as a society, need to take to arrive at a desired future and avoid an undesired future.

2. Vignette

In 2025, artificial intelligence continues its technological march through many applications in all walks of life. Players of massive multiplayer online games use avatars of themselves or their movie heroes.[37] If a dragon scorches an avatar, no problem. The player can easily create

Do you think the use of a vignette helps to makes it easier for stakeholders to relate

[35] Signal Processing, "How to mimic/copy/fake someone's voice?". https://dsp.stackexchange.com/questions/7833/how-to-mimic-copy-fake-someones-voice.

[36] Vincent, James, "Lyrebird claims it can recreate any voice using just one minute of sample audio", *The Verge*, 24 Apr 2017.

[37] One commentator has observed that "there are a lot of people out there obsessing about creating a way for users to generate, personalize and 'own' their avatars... Loom's technology turns selfies into personalized 3D avatars by applying machine learning to automate human face visualization. It uses public APIs and VFX to create life-like visualizations which can then be animated and used for a range of applications. Video embedded above shows how an avatar generated from a single inset image (in these cases of celebrities such as Will Smith and Angelina Jolie) can look remarkably life-like and expressive." Bonasio, Alice, "What does the Future of Avatars Look Like?", *Tech Trends*, 13 Dec 2016. The Dutch government is funding research for the development of an avatar of oneself for medical purposes. "A medical avatar is a virtual copy of yourself, on which you can read your own personal data and with which you can monitor your health well." See: Zorgvisie [NL], "Nederlandse onderzoekers strijden om 1 miljard euro voor virtuele avatars", [Dutch researchers are competing for 1 billion euros for virtual avatars], 20 Sept 2018. https://www.zorgvisie.nl/nederlandse-onderzoekers-strijden-om-1-miljard-euro-voor-virtuele-avatars/?tid=TIDP207360XD9A308C89FC4492EA729DB8920717470YI4&utm_medium=email&utm_source=20180920%20Zorgvisie%20nieuwsbrief%20-%20dagelijks&utm_campaign=NB_Zorgvisie

another avatar who looks and behaves exactly like the first one. Criminals use the same technologies to mimic a target's friend or relative who is in urgent need of funds because someone stole their purse in Chicago.[38] Brad Pitt and Scarlett Johansson have been distressed to find their faces and voices used in porn films.[39] Politicians are accused of spouting incendiary statements they did not actually make.[40] In 2025, technologies that mimic people are being used to create companions for senior citizens. With the ageing population, governments are finding it more of a challenge to provide social services and assisted living facilities to all those in need. Hence, some governments began investigating the possibility of using artificial intelligence and a set of other technologies in social care applications, both as a cost reduction measure, and as a way of overcoming the shortages of trained staff.[41] Some activists feel that senior citizens should have the right to have a real human, rather than a machine, as a carer, but the cost of personalised holograms is dropping at a time when it is difficult to recruit enough human carers, doctors and nurses to take care of our ageing population. A public consultation in 2024 showed that a majority of respondents favoured the deployment of holographic support services. Although the research is still preliminary in 2025, sociologists and physicians are in general agreement that senior citizens who engage with their holograms or personalised avatars are likely to live longer. Alfred's wife of 45 years died in 2024. He missed her greatly until a government agency told him that he could have a hologram who could interact with him just like his dear Lucy. The hologram knows about their lives together. AI has synthesised all of Lucy's data, is able to	to the technology and its impacts?

[38] Welser, William, "Fake news 2.0: AI will soon be able to mimic any human voice", *Wired*, 8 Jan 2018.

[39] "Advanced machine learning technology is being used to create fake pornography featuring real actors and pop stars, pasting their faces over existing performers in explicit movies." Hern, Alex, "I used to face-swap Hollywood stars into pornography films", *The Guardian*, 25 Jan 2018.

[40] Hsu, Jeremy, "Experts Bet on First Deepfakes Political Scandal", *IEEE Spectrum*, 22 June 2018.

[41] Steps are already being taken to develop holographic companions. Japanese company Vinclu has developed "its Gatebox virtual assistant, which features a holographic anime character that can provide companionship to lonely individuals. First introduced as a concept 'communication robot' in January [2016], the new holographic virtual assistant goes by the name Azuma Hikari and is given the appearance of a female anime character." F., Jessie, "Gatebox Virtual Assistant Is The Holographic Anime Companion For Lonely People: Can It Compete With Amazon Echo?", *Tech Times*, 19 Dec 2016. See also Vincent, Brittany, "Gatebox Wants To Be Your Personal Holographic Companion", Geek.com, 16 Dec 2016: "The device, seen here with pre-installed character Azuma Kikari, can be used to automate things like your lights at home and other processes as well as act as a miniature companion in your home. The home robot is seen acting as both and fulfilling several functions in the touching video advertisement, where a lonely salaryman texts and receives texts from the robot during the day, interacts with her as if she were his wife, and enjoys her company." See also Fresh Technology, "Holographic Home Companions: Can AI Technology Cure Loneliness?", 2 May 2017. See also a comment by a blogger: "As for the looks of the hologram, he programmed it to look like his favorite person, Veronica". https://taleoftwowastelands.com/viewtopic.php?p=41190#p41190. The Spike Jonze 2013 film *Her*, starring Scarlett Johansson, already foresaw a computerised companion.

reproduce her voice[42], her appearance, her mannerisms, even the way she used to argue with him.[43]

The creation of Lucy the hologram was made easier because Alfred and Lucy had been using home assistants, like Siri and Alexa, for many years. Although Alfred was initially a bit wary of this new Lucy, social services convinced him that he would live longer and be happier with this Lucy and be less dependent on social services. It did not take him long to accept this Lucy as soon as she reminded him to take his daily meds and to go for a walk because he needed the exercise.[44]

Not only has Lucy the hologram absorbed the data that belonged to her predecessor, she keeps up to date with the wearables that monitor Alfred's health and well-being as well as the sensors that monitor the status of various appliances and processes (heating, water, power) in his home.

3. Drivers

Several key drivers have impelled the development of technologies and applications that mimic people. Among them are the following:

The adult entertainment industry is often at the forefront of visual technological advances (e.g., online video streaming, virtual reality etc). As mentioned in the vignette, it is likely that this industry will be a key driver of the adoption of this type of technology.

Following economic and social studies, governments became convinced that technologies like those behind Lucy help them respond to the needs and demands of an ageing population.

Technologies like Lucy provide safety and security for their owners, by being aware of the owner's ambient environment and how they should respond to events. While holograms cannot stop intruders, robots will be able to provide such protection. Holograms or avatars like Lucy can, however, provide a link with the world outside Alfred's home and can alert the police or a doctor should the need arise. Buyers of such technologies need to consider the pros & cons of robots versus holograms.

Competition has been an important driver in the development of technologies that mimic people. Researchers and scientists in several countries have been working on the same technological capabilities.

Do you agree with these drivers?

Are there any other significant drivers that should be included here?

[42] Metz, Rachel, "Google demos Duplex, its AI that sounds exactly like a very weird, nice human", *MIT Technology Review*, 27 June 2018.

[43] IBM's "Project Debater uses conversational interfaces to debate any topics like humans, rationalizing arguments and even showing a bit of sarcasm and a sense of humor." See Rodriguez, Jesus, "The Artificial Intelligence Research Behind the Impressive Project Debater", Medium, 20 June 2018.

[44] There are reports that some players of "dating sims" fall in love with digital characters. "The most dedicated romantic gamers do not see their interactions with virtual characters as a substitute for human companionship, but as a new type of digital intimacy." Schwartz, Oscar, "Love in the time of AI: meet the people falling for scripted robots", *The Guardian*, 26 Sep 2018.

Amazon, Apple, Facebook, Google and Microsoft ("the big five") have all seen huge market potential in the development of Lucy-like holograms, avatars and robots. Other countries with ageing populations such as Japan and South Korea are leading the way in the development of holographic companions. In addition to competition at the geographic and institutional level, competition exists between those who favour open source and those who favour proprietary technologies.

Research & innovation factors – Funding from DARPA in the US and the European Commission has contributed to the development of the technologies, especially by small and medium size enterprises (SMEs) and universities. The big five have been developing and patenting such technologies without government support. Meanwhile, the smart home has become a reality with dozens of connected technologies, from smart home locks and thermostats to lights to sensors capable of detecting things like falls. The need for a centralised point of communication with these tools spurred a big part of the rise of second-generation AI assistants, like Alexa and Google Home.

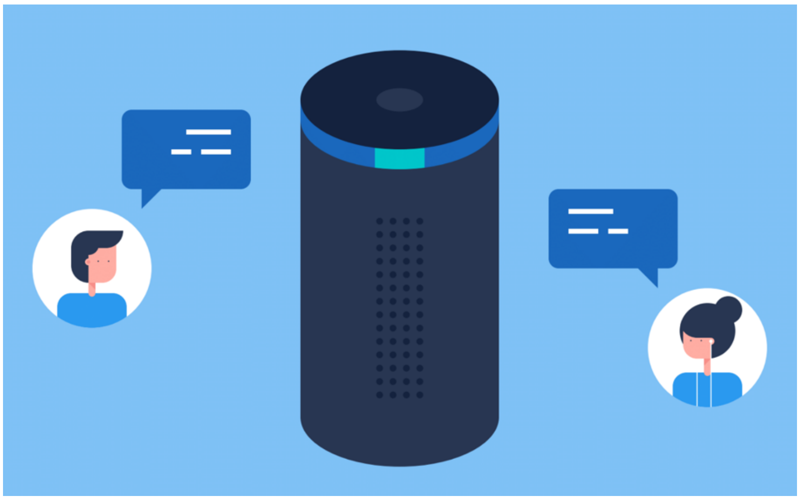


Figure 3 AI assistants - Image Credits: Tony Webster, Flickr (CC BY 2.0)

A shortage of carers for senior citizens – Civil society organisations, such as Age UK, have long pressed for more support for senior citizens. With the rapid increase in the numbers of senior citizens, governments are pushed to provide human care for all those who need it. In addition, the voting power of senior citizens has convinced governments of the need to support artificial carers.[45]

Data availability – has been an important driver in the development of Lucy. Although the EU's General Data Protection Regulation (GDPR) has made organisations more sensitive to the public release of personal data, new technologies, such as the Internet of Things, have greatly expanded the ready availability of data that Lucy the hologram needs to be credible to her owner.

[45] http://www.age-platform.eu/

The cost of supporting senior citizens has ballooned past the resources of most national governments. The ageing population needs care and support. Studies of the needs of senior citizens have shown that technologies that mimic recently deceased spouses diminish the demands on social services and health services, which suffer from shortages of doctors and nurses.	
4. Barriers and inhibitors Despite the potential of an ageing population, the actual market size is uncertain. The cost of Lucy technologies is dropping fast. Lucy and her peers are still beyond the means of most people in 2025, although market projections suggest that in the next five years such technologies will be commonplace. However, related to issues of cost as a barrier, another recession will likely slow down development of these technologies. Austerity might force governments and other organisations to curtail their research funding and delay introduction of such technologies. Training mimicking technology is a not insignificant task, especially where the technology is to serve as a companion to senior citizens. Home assistants help when both seniors are still alive, so that they can record increasing volumes of data about the senior citizens in whose home they occupy a critical space. Home assistants are doubly useful – not only to capture data, but also to put users at ease with the technology. Production of a Lucy requires massive amounts of data, especially personal data. It is possible that there will be a 'Facebook/Cambridge Analytica' type of incident that may temporarily set back development of this type of technology. There could be a data breach, a misuse of data or discovery of unwanted invisible uses of data. This sort of "avatar" poses other security concerns: impersonation (for malicious purposes, such as data exfiltration, i.e., stealing the senior person's personal data, credit card numbers,…), secure data storage (the hologram will have access to a lot of data, potentially visual records of the senior person accessing its personal information, such as bank account, and might even help him/her do so, and thus have full access to such data), ethical data mining and machine learning (how does the company that provides this service make sure it does not pool the data from numerous users to create the hologram, and thus "leaks" other people's information into the service they provide). This could lead to a public backlash and boycott of the technology. Whether data breaches, even massive ones., will be concerning enough for the general public to worry and boycott the technology accordingly, is difficult to predict. Recent examples of data breaches show how little the general public is concerned about breaches affecting other people, which suggests such breaches have become the norm.	Do you agree that these are likely to be the most significant barriers and inhibitors in 2025? Are there any other barriers that should be mentioned?

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

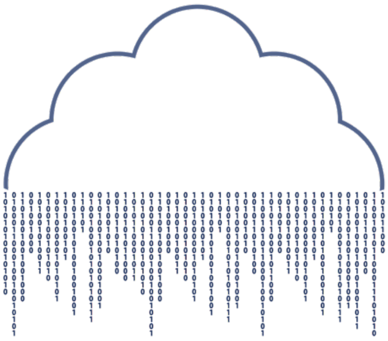


Figure 4 Data breaches

There are issues surrounding the use of such data, not just in the regulatory context of data protection, but also in the context of a philosophical question: Will Alfred accept an AI technology that is much "smarter" than him, that can recall events he has forgotten, that can explain how things work, that can guide him in taking better care of himself? Also, correlating data from many different sources has been a technical challenge for the past decade or so, but tremendous progress has been made in correlating, synthesising and interpreting data, which have been prerequisites to create holograms like Lucy. In 2025, there has been an ongoing debate about deepfake avatars, and which are better for senior citizens; holograms, avatars or robots? A robot that could mimic Lucy, that could look like her and act like her is more challenging than a hologram. By 2025, there has been huge progress in creating humanoid robots, but even so, they are considerably more expensive than holograms, but they have the advantage of being able to move things in the physical world. Hence, they can perform housework and other tasks, such as helping to protect their owners against intruders.

The diffusion of these technologies may be inhibited if an incident is made public where domestic robots have forced people to do something unfavourable, e.g., to eat or drink something the owner didn't want or where the robot was trying to convince its "owner" to do something that clearly was not in the owner's interest.

Another potential barrier is the availability of bandwidth for remote presentation of holograms. As an indicator of potential bandwidth requirements, the throughput to run virtual reality is almost 100 times higher than a high definition video.[46]

Trust is an inhibitor to the adoption of Lucy-like holograms and avatars – senior citizens need to trust them. Stories in the press about holograms that behave erratically don't help. There is a general consensus that machines should periodically remind their "owners" that they are machines, but others ask: what is the point of having a technology that mimics people saying that it is a machine?

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

5. Ethical, legal, social and economic impacts

In 2025, the benefits of technologies that mimic people are especially apparent in support of senior citizens. The benefits are not, however, unalloyed, as the following paragraphs point out.

Ethical impacts

While the high cost of human care has led to development of home-care holograms like Lucy, still some ethicists and other stakeholders question the legitimacy of substituting a technology for a human carer. Still others question the ethics of “reincarnating” a deceased spouse as a futuristic Alexa assistant. By giving these devices a human appearance, a line is more thoroughly drawn in terms of making them human replacements: something that has profound application when, for instance, using it to carry out medical diagnosis. Is Alfred more likely to follow Lucy’s diagnosis than a similar injunction on a computer screen? No one knows, but it is a subject of research in 2025.

Do you agree that the ethical issues listed here are likely to be important in 2025?

Are there any other ethical issues we should include?

The interaction of Lucy, Alfred’s wearables and the sensors in his home also raises complex ethical issues about autonomy, equity and sustainability. For example, have the various technologies stripped away Alfred’s ability to function as an autonomous individual? Alfred is privileged because he has Lucy, but it seems likely that he is developing a dependency on his new Lucy. He is one of the few members of the public to have a personalised hologram, which raises issues of social equity.

The holograms raise issues of sustainability too. When Alfred dies, what happens to Lucy? Is she simply switched off and allowed to die a digital death? Will anyone miss all of the knowledge that Lucy has acquired, not only of her human predecessor, but also of Alfred?

Some governments insist on taking partial control of Lucy and her peers. In some instances, the partial control is for Alfred’s own safety and well-being. Lucy can prompt Alfred to take his medicines and encourage him to do some physical exercise or to converse with her instead of watching TV all day long. But in other instances, governments prompt Lucy to quiz Alfred about whether he is working part-time or has some other source of income so that government can reduce its benefits to Alfred. So an ethical issue has arisen as to whether technologies that mimic people should be totally controlled by the “user” (by Alfred) or whether control over Lucy should be shared with governments or the company that has created Lucy. Who controls Lucy raises a question of free will for Alfred. If he does not want to take his medicine, and Lucy wants him to, what happens? Does he get forced to do so (psychologically, physically)?

[46] https://www.manchestereveningnews.co.uk/news/greater-manchester-news/vodafone-5g-network-trial-manchester-15178160

Many privacy advocates continue to express concerns that the holograms, avatars or care robots are actually sophisticated surveillance agents as they pass on the information they collect about their owners to the big tech companies and government agencies. Because of such allegations, some governments have established ethical committees to advise on issues raised by AI. Some have called for a global agreement to govern ethical issues raised by technologies that mimic people.

There have also been concerns about whether holograms, like Lucy, can make medical diagnoses. Studies have shown that holographic people are more often right in their diagnoses than real doctors. While Lucy is probably perfectly capable of making a much better diagnosis and prescription than a real doctor, who is responsible in case of a mistake? Who gets blamed?

By giving these devices a human appearance, a line is more thoroughly drawn in terms of making them human replacements: something that has profound application when, for instance, using it to carry out medical diagnosis. Does the use of a human-like avatar suggest human-level success at tasks? Does it raise the possibility of human-level failures in events that machines may be able to perform better?

If Lucy is “smarter” (more knowledge-informed) than Alfred, will she always (agree to) be subservient to Alfred? If the holograms like Lucy or robots reach levels of intelligence close to pets or animals or humans, what is the acceptable threshold where it is no longer a robot or a machine that one can freely discard and use or abuse eventually as a tool, and it becomes an entity that should be recognised as a being with rights?

Among other ethical issues being discussed in 2025 are the following:

- Human rights issues such as self-determination. Lucy may be able to manipulate Alfred in various ways, to enforce a routine that may be too restrictive, to induce him to buy certain products or services, or to offer criteria he should consider in deciding his vote.
- Should robots have a legal personality? Should they show respect personally and culturally to human beings?
- How do holograms like Lucy impact our privacy and right to be forgotten? Did Lucy’s human predecessor agree to be replicated by a hologram?
- What enforcement powers should regulators have against instances of manipulation?
- What are optimal mixes between self- and co-regulation and legislation and public supervision of the technologies that mimic people? Should there be restrictions on who can be mimicked? What are the transnational aspects of these technologies?
- How can we embed the precautionary principle in innovations such as Lucy?

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

- Who should have access to all known data about Lucy? Who should have access to the data that Lucy collects from and about Alfred? If Alfred has full control of Lucy and all her data, could he renounce such control in order to obtain a better insurance policy? If Alfred were still employed and his employer wanted access to the data Lucy has acquired, would he feel obliged to give it to him or her? If Alfred's doctor performed a diagnosis, who would be given the results? Alfred? Lucy? The health care system? [47]	
Legal impacts In 2025, some legal issues surrounding such technologies have been resolved, yet others are still being debated. **Transparency** is one such issue. Lucy's designers have programmed her to explain why she does something when asked. Transparency is also an issue with regard to data sources. Alfred might ask 'How does Lucy know so much about me?' and the answer to that is a data protection transparency issue. If Lucy malfunctions, **who is liable**? [48] Is it the company that created Lucy? Is it Alfred? (The company could claim that he sent contradictory commands that confused poor Lucy.) Is it the designer, the manufacturer, the programmer, the trainer, the data provider? What is the threshold for a causal link in case of damages? Such questions plague the courts in 2025. Closely related to liability, **accountability** is critical to ensure that AI algorithms perform as expected. Finally, in 2025, the European Court of Justice has ruled that it is not sufficient to hold humans accountable for the actions of the AI algorithms they create, but that the concept is more nuanced, i.e., AI systems need to explain and justify decisions and actions to Alfred and others with whom Lucy interacts.[49] The issue of a company secretly using data for the secondary purpose of advertising (**invisible processing**) is also a legal issue. Data protection	Do you think these will be the key legal issues in 2025? Are there any other legal issues that we should include?

[47] Mantelero, Alessandro, "AI and Big Data: A blueprint for a human rights, social and ethical impact assessment", *Computer Law & Security Review*, Vol. 34, 2018, pp. 754–772.

[48] Microsoft is calling for government regulation of facial-recognition software, one of its key technologies, saying such artificial intelligence is too important and potentially dangerous for tech giants to police themselves. Harwell, Drew, "Microsoft calls for regulation of facial recognition, saying it's too risky to leave to tech industry alone", *The Washington Post*, 13 July 2018.

[49] See Dignum, Virginia, Matteo Baldoni, Cristina Barolglio, et al., "Ethics by Design: necessity or curse", Paper 68 presented at the AAAI/ACM conference on Artificial Intelligence, Ethics and Society, New Orleans, 1-3 Feb 2018. "Accountability refers to a system's need to explain and justify decisions and actions to its partners, users and others with whom it interacts. To ensure accountability, decisions must be derivable from, and explained by, the decision-making algorithms used. This includes the need for representation of the moral values and societal norms holding in the context of operation, which the agent uses for deliberation. Accountability in AI requires both functionality for guiding action (by forming beliefs and making decisions), and for explanation (by placing decisions in a broader context and by classifying them along moral values."

legislation (e.g., GDPR) prohibits further uses of data that are incompatible with its original use, e.g., a social care AI system that uses the profile it builds up on its subject to nudge them towards purchases. Such secondary use breaches the fairness, transparency and lawfulness principles.	
Social impacts Some sociologists and psychologists and psychologists have expressed concern that Lucy can create dependencies, much like home assistants. Users such as Alfred tend to respond to Lucy-type creations in several different ways. Some users are bemused by the technologies. They continue to recognise that the holograms and robots are machine-created, are not the real thing, can never replace the real thing, but the hologram is a noble effort nonetheless in attempting to recreate a loved one. Other users become psychologically attached to the holograms, avatars or robots, much as they form attachments to pets. They treat the holograms as the real thing ("Do you remember our trip to Wyoming?"). Still others reject the holograms in irrational ways: They taunt the avatars for not correctly "recalling" a shared event. In 2025, with the increasingly lifelike holograms, avatars and robots, experts and many senior citizens continue to debate the rights of such creations. Experts and ethicists thought they had dispensed with this issue in 2020, but it has returned in 2025, in part, because Lucy-type robots are expensive. Users, designers, programmers, manufacturers and social services all wish to protect their investments and what better way than attributing rights to Lucy[50], e.g., the right to dignity, the right to integrity of the person, the right to security, freedom from non-discrimination, freedom of expression and information, the presumption of innocence and even the right to good administration. Robots and holograms gaining rights could give rise to further issues in the balancing of rights against those of natural individuals.[51] For instance, in data protection terms, a natural individual may wish to exercise their right to erasure (the right to be forgotten). If exercised, this would require the deletion of much of the data used by the hologram/robot, impinging on its own privacy rights in relation to personality, self-development, etc. Although the cost of personalised holograms and robots is dropping, some social tension has arisen with claims that such creations are only affordable by rich people, that they widen the gap between rich people and the rest of society. For those who wish to have an AI mimicking a	Do you think these are likely to be important social impacts created by these technologies in 2025? Are there any others you think we ought to include?

[50] Attributing rights to a hologram may not be so far-fetched. At least one government has granted rights to a robot. Wootson Jr., Cleve R., "Saudi Arabia, which denies women equal rights, makes a robot a citizen", *The Washington Post*, 29 Oct 2017.

[51] Yampolskiy, Roman V., "Could an artificial intelligence be considered a person under the law?", *Phys.Org*, 5 Oct 2018. https://phys.org/news/2018-10-artificial-intelligence-person-law.html

loved one, there is a requirement to have lots of data to train the underlying algorithms. People who historically have had limited access to the types of technologies that harvest this data (e.g., due to cost or disabilities) may not have the requisite data, and therefore may not be able to take advantage of such products. Where AI mimicry is of poor quality (limited training data), there may be poor quality outcomes for people due to imprecise predictions and decision-making by the AI. An 'off-the-shelf' mimic may be sold as being 90% accurate at mimicking loved ones. But this could be 100% accuracy for 90% of the population (e.g., white westerners) and 0% accuracy for 10% of the population (e.g., minorities).	
Security and economic impacts While technologies that mimic people have created thousands of new jobs in 2025, they have also created new opportunities for malefactors (criminals, terrorists) who have hacked some holograms and robots so that they won't respond to voice commands until a ransom has been paid. There are various sources of the hacking of computer-generated companions like Lucy, including individuals (pranksters, trolls), organised crime (extortion, blackmail, fraud), and government (surveillance, spying). A particular concern is adversarial learning, in which the learning mechanisms of algorithms can be misled and can cause AI systems to make bizarre and unpredictable decisions.[52] Amazon produced Lucy the hologram. Sometimes, Lucy suggests that Alfred consider buying something that improves the quality of his life. At other times, Lucy tries to convince Alfred to buy stuff he doesn't need. She does it in a subtle way so that it is not obvious to Alfred that Amazon is manipulating him. Activists and civil society organisations have railed against such practices. Government studies and research by several think tanks have shown that the use and deployment of holograms, avatars and robots have an overall positive economic impact. They reduce the need for providing healthcare and other social services, because Lucy and her cousins can give medical and healthcare advice to their "owners" (a term in much social contention). Their development, deployment and ongoing research create high-quality, high-value jobs. Indeed, there is much demand for trainers, those who "train" the machine-learning algorithms with every scrap of data that ever existed about Lucy and her peers so that Lucy the hologram appears to know more about dead Lucy than Alfred. The source of the data necessary to train the algorithms will often be the private sector (e.g., creators of home assistants and IoT devices).	Do you agree with these security and economic impacts? Are there any other security and economic impacts that you think will be particularly important in 2025?

[52] Buchanan, Ben, Prepared Testimony before the House Oversight Committee, Belfer Center for Science and International Affairs, Harvard Kennedy School, 18 Apr 2018. https://www.belfercenter.org/publication/prepared-testimony-house-oversight-committee

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

There are ethical questions around the role that companies with proprietary products have to play in the provision of social care and healthcare in countries such as the UK and Canada with publicly owned healthcare systems. Private firms are first and foremost answerable to their shareholders, not their customers. With the advent of avatar companions, today's incumbent tech firms may become even more powerful and see their data monopolies strengthened by 2025.

Proponents of data-driven technologies often argue that it will create new jobs and new skill sets for existing workers. At the same time, critics argue that these technologies are actually resulting in job losses in domains where automated (or autonomous) processes render human involvement redundant in 2025. AI mimics are replacing large swathes of social care jobs while the public and private sectors are not doing much to re-skill displaced workers.

6. Mitigating the negative and acting on the positive impacts

By 2025, various countries have taken different actions to mitigate the negative impacts of AI in social care and to accentuate the positive impacts.

Do you think these actions are plausible and probable?

Several countries – principally, the US and Canada -- and the EU have established AI advisory councils. There were (and still are) numerous calls to establish regulators with enforcement powers ("with teeth"). However, industry, some politicians and other stakeholders have argued that because of the rapid increase in AI applications and uncertainties about their impacts, regulatory action could severely retard innovation. In the end, the EU decided to create an AI advisory council (somewhat modelled after the European Data Protection Board) as a first step, with formation of a regulator as the envisaged next step, if necessary.

Some have argued that the advisory council approach is a sop to industry, that industry can ignore such advice and do as it likes. However, such has not been the case. Employees in the big five have become activists. Increasingly, they seek to implement Google's original dictum – "Do no evil!". As long ago as 2018, Google employees pressured the company to ban development of AI software in weapons. The company also established strict ethical guidelines for how the company should conduct business in the age of AI.[53] Similarly, Microsoft employees sent a letter of disapproval to their chairman about the dangers of facial recognition technology, which led to calls for regulating the technology.[54]

[53] Harwell, Drew, "Google bans development of artificial intelligence used in weaponry", *The Washington Post*, 7 June 2018.

[54] In a letter to their chief executive, Microsoft workers said they "refuse to be complicit" and called on the company to "put children and families above profits." Harwell, Drew, "Microsoft calls for regulation of facial recognition, saying it's too risky to leave to tech industry alone", *The Washington Post*, 13 July 2018.

While governments dithered on the issue of regulation, employees of the big five have been the prime movers in the adoption of an ethical approach to AI. Ethics by design has become as commonly espoused as privacy by design. Nevertheless, civil society organisations have expressed scepticism about the "real" intentions of the big five, alleging that their ethical initiatives are simply charades to put governments off regulation.

Employee activism has precipitated a wave of ethical introspection among the big five as well as many other smaller companies. Given the sensitivity of health and social care, industry created ethical advisory councils to review the development and deployment of holograms, like Lucy, as well as robots, not only those used as companions for senior citizens, but those used in other domains such the porn industry, political campaigns and targeted advertising.

Governments have not been totally lame. Governments have created new offences (use of technologies that mimic people without consent of the person mimicked), punishable by fines and prison sentences of up to two years. Governments have promoted AI standards and public awareness of the issues raised by technologies that mimic people.

In addition to the actions described in this section, standards bodies are playing a role in mitigating the risks of these technologies. For example, IEEE's P7000 series of standards was developed to address the ethical issues in the design of AI and autonomous systems.[55] Governments may need to assess how effective these standards are.

7. Recommendations for a desired future and avoiding an undesired future

Our desired future is one where technologies that mimic people are strictly controlled to beneficial applications, like Lucy. Regulators with enforcement powers are deemed necessary if society is to avoid an undesired future where there are no controls over how such technologies are used, whether for healthcare applications, political manipulation, pornography, fake news, etc.

Just as there are different stakeholders in the use of technologies that mimic people, so there are different steps that stakeholders can take toward a desired future. Among the conclusions and recommendations of those who contributed to this scenario are the following:

- Academics should explore the ramifications of the new technologies and, where possible, ensure technologies that mimic people are open source. Before holograms or robots are used in social care applications, such as that depicted in the vignette, developers and/or operators should conduct a data protection impact assessment.

Do you agree with these recommendations?

Are there any others that you think we should include?

[55] https://standards.ieee.org/industry-connections/ec/autonomous-systems.html

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

● Industry should develop and use ethics councils within individual companies and as well as across companies. Such councils should be multi-disciplinary with people from backgrounds such as legal, risk, compliance, data science, software development, design, user experience and ethics. Industry stakeholders should come together to create a road map for the development of such technologies and a set of principles to govern their use. ● Policymakers should initiate public consultations about regulatory options governing mimicking technologies, especially where they are used to perform social care functions. ● Regulators should use their enforcement powers proportionately. They should find novel ways to work with industry to support compliant and ethical innovation in AI mimicry. ● Sector regulators and industry bodies should create codes of conduct for the use of AI mimicry in particular contexts, including social and health care. ● Existing regulators should adopt a co-ordinated (co-regulatory) approach to AI mimicry to ensure harmonised, consistent rules for industry. As holograms like Lucy raise various issues beyond the remit of a single regulator, some mechanism is needed to ensure regulatory harmonisation. ● Governments should support secure, compliant access to representative datasets for training purposes. This should help ensure higher quality offerings for traditionally under-represented parts of the population, while also addressing issues around data monopolies by giving SMEs access to training data. These data may include biases or influence certain opinions or actions. Furthermore, competition issues around data cannot be solved exclusively by governments providing training data. ● Governments should embed ethics and compliance into the curriculum, and in particular higher and further education courses in subjects such as computer science, so that data scientists are exposed to scenarios such as those in this deliverable. ● Governments should support training programmes for workers likely to be displaced by AI mimicry. In the scenario, Lucy may displace human social care workers. ● Governments and transnational companies, including the big five, should begin work on a global agreement on the legitimate and unethical uses of such technologies – like a Wassenaar Arrangement on AI.[56] ● Governments and industry should encourage artists, directors, film producers, to create TV films[57] or films showing positive and negative side of technologies that mimic people.	

[56] The Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies is a multilateral export control regime. https://www.wassenaar.org/

[57] Such as the UK Channel 4 production *Humans* and HBO's *Westworld*.

• Governments should encourage shareholders’ participation in major decisions about AI. Policy scenarios are one important way to do so. • The fundamental question should not be: what can we do with new technologies, but how can new technologies help individuals on their own terms and convince them that new technologies are ethical and promote equality, well-being and trust in democratic values?	

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

4. Second scenario: Information warfare in 2025

Figure 5 Second scenario

1. Introduction

This scenario considers the nature of warfare in 2025 and, in particular, cyber warfare or information warfare. The scenario makes the point that the nature of warfare has changed, and government research programs can no longer afford a blurred boundary between civilian and military research. The scenario argues that in many instances it is not known who is behind a cyber attack. In other cases, where certainty exists, retaliation seems warranted. In the following paragraphs, we describe how the nature of warfare has changed. We note some of the new technologies and applications before we get to a vignette that casts doubt on who is behind a cyber attack on a nuclear power plant. We then discuss the drivers behind information warfare, its impacts and our recommendations to policymakers who have to deal with the continuing high cost and travails of information warfare.

So now let’s start with the situation in 2025…

The weaponry of war continues to evolve – from bows and arrows to nuclear bombs to algorithms. The field of battle has changed too – from the physical world to an invisible world, but no less dangerous for that, with real-world consequences.

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Nuclear war has been avoided so far, because each side knew that nuclear war would be the end of civilisation. In cyber war, there is no similar principle of mutually assured destruction (MAD) to avert disaster(s). Until recently, we knew who the warring states were. In cyber war, such certainty is much more difficult. Cyber attackers can easily cover their online tracks.[58] The nature of warfare is changing. It has become a global phenomenon in 2025. It involves many different actors, from governments to cyber gangs.

Attacking an adversary no longer requires massive bombing runs or reams of propaganda. All it takes is a smartphone and some software readily available on the dark web. There have already been many cyberattacks in recent years sponsored by states or their subsidiaries. The frequency of cyberattacks is increasing.[59] Politicians are calling for stronger action against cyberattacks.[60] The nature of cyberattacks is also changing. Attackers are no longer attacking just critical infrastructure; they are also attacking whole populations, trying to sow disruption of public opinion and electoral processes. As a result, governments and businesses are increasing their budgets for research on how to contend with the increasing frequency of cyberattacks and the huge risks that come when one country uses smart information systems to disable another country's critical infrastructures and social consensus.[61]

Warfare technologies in 2025

What does artificial intelligence in warfare mean? AI can be used for offensive and/or defensive purposes; it can take many forms, but essentially AI comprises algorithms capable of processing and learning from vast amounts of data and then taking decisions autonomously or semi-autonomously. In 2025, AI is used in many weapons systems, in the tangible world as well as in cyber space. Here are some material examples.

[58] "To mask themselves, attackers generally compromise computer servers and networks operated by other organizations. These nodes then serve as unwitting springboards for further electronic assaults." Hackett, Robert, "How Hackers Plan Attacks and Hide Their Tracks", *Fortune*, 12 Aug 2016.

[59] Graham, Luke, "The number of devastating cyberattacks is surging — and it's likely to get much worse", CNBC, 20 Sept 2017. See also: "Cyberattacks are becoming more frequent, sophisticated and destructive. Each day in 2017, the United States suffered, on average, more than 4,000 ransomware attacks, which encrypt computer files until the owner pays to release them. In 2015, the daily average was just 1,000." Taddeo, Mariarosaria, and Luciano Floridi, "Regulate artificial intelligence to avert cyber arms race", *Nature*, Vol. 556, 19 Apr 2018, pp. 296-298. "Global damages from cyberattacks totalled $5 billion in 2017 and may reach $6 trillion a year by 2021".

[60] See, for example, Stewart, Heather, and Jennifer Rankin, "Theresa May to urge EU leaders to take action on cyber-attacks", *The Guardian*, 17 Oct 2018. See also Press Association, "UK 'wholly' unprepared to stop devastating cyber-attack, MPs warn", *The Guardian*, 19 Nov 2018.

[61] Morgan, Steve, "Worldwide Cybersecurity Spending Increasing To $170 Billion By 2020", *Fortune*, 9 Mar 2016. Such issues also feature in recent publications, such as Moore, Martin, *Democracy Hacked: Political Turmoil and Information Warfare in the Digital Age*, Oneworld Publications, 2018, and Singer, P.W., and Emerson T Brooking, *LikeWar: The Weaponization of Social Media*, Eamon Dolan/Houghton Mifflin Harcourt, Oct 2018.

- Tactical battlespace development – AI is used in autonomous vehicles in reconnaissance and offensive and defensive roles, e.g., killer vehicles, access denial systems like smart mines or automated systems that respond to attack.
- Sensor fusion – AI is used to bring information from many sources – e.g., satellites, aerial reconnaissance and local information feeds to and from the war fighter – and develop a coherent view of threats and potential threats faster and more accurately than painstaking analysis by trained analysts.
- High-speed, high-intensity warfare – AI systems identify potential threats and launch countermeasures at high speed since humans are unable to respond fast enough. Machine decision making is required to take action to defeat a potential threat.
- Target identification – AI drives facial recognition systems used to identify individuals, and even ethnic groups, for potential consequences.
- Making sense of non-conventional warfare or terrorism – The emergence of driverless cars in the last few years has created another potential to carry out devastating attacks without the risk to the terrorist's life. The emergence of drones, potentially piloted remotely or pre-programmed to fly to a given location, add yet another layer of threat. As someone put it – in every garage, a bomb.

In 2025, many states use artificial intelligence in cyberattacks and cyber defences. These states include the US, UK, Israel, Russia, China, Iran, North Korea, among many others. The states have been investing billions in AI-enabled cyberattack and cyber defence systems.

Powered by AI, cyberattacks occur more rapidly and widely. As governments and companies have learned the hard way, they need to invest in cyber defences, in making their organisations more resilient and in raising public awareness of the risks of being manipulated. Governments and companies are spending billions of euro in 2025 to increase their cyber expertise and defences.

Attackers use smart information systems (SIS) that combine AI and big data in multiple ways. They use AI-powered bots to game the algorithms used in other systems.[62] They use driverless cars as bomb delivery vehicles as an alternative to human suicide bombers. They use AI in stealth weapons and in pattern recognition and in deepfake technologies. The latter are particularly insidious as it becomes impossible for ordinary citizens to know whether they are being fed facts or fabrications.

Autonomous weapons systems, including killer drones, drone swarms, robot soldiers, submarines and tanks without a crew, have been

[62] SafeGuardCyber, How Russian Twitter BOTs weaponize social Media to influence & DISINFORM, p. 10.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

developed in many countries, including the China, Israel, Russia, South Korea, the United Kingdom and the United States.[63] AI powers many routine tasks, which has improved mission times and the precision of attacks on targets.[64] In 2025, warfare is fought by highly intelligent, inscrutable algorithms that speak convincingly of things that never happened, producing "proof" that doesn't really exist.[65]

AI is used to automatically create personalised phishing e-mails for social engineering attacks to target thousands of people at the same time. AI is used to mutate malware and ransomware more easily, and to search for and exploit vulnerabilities in all kinds of systems. AI produces complex and highly targeted scripts at a rate and level of sophistication far beyond any individual human hacker.

Figure 6 Phishing

Cyber defenders also use AI to process large volumes of data to help detect attacks against critical infrastructures. The big social media companies claim they identify millions of fraudulent or malicious accounts (liars) per day.[66] In addition to detection technologies, AI is used in forensics and fault diagnostics. In 2025, soldiers interpret information faster and more quickly recognise threats like a vehicle-borne improvised explosive device, or potential danger zones from

[63] Holley, Peter, "Tech leaders: Killer robots would be 'dangerously destabilizing' force in the world", The Washington Post, 19 July 2018.

[64] Szondy, David, "Israel unveils Merkava Mk 4 Barak smart tank", New Atlas, 22 July 2018. https://newatlas.com/merkava-mk-4-barak-smart-tank/55556/

[65] Singer, P.W., and Emerson T. Brooking, "The Machines That Will Fight the Social Media Wars of Tomorrow", Gizmodo, 5 Oct 2018. https://gizmodo.com/the-machines-that-will-fight-the-social-media-wars-of-t-1829445747

[66] "Facebook believes artificial intelligence can be used to support its efforts to identify bad actors... The company has said its technology can block millions of accounts a day as they are being created, before they spread fake news or inauthentic ads." Zakrzewski, Cat, "Technology giants face big test in midterm elections", *The Washington Post*, 8 Oct 2018. https://www.washingtonpost.com/news/powerpost/paloma/daily-202/2018/10/08/daily-202-technology-giants-face-big-test-in-midterm-elections/5bba75ea1b326b7c8a8d1886/?utm_term=.3c1c7e90523e

aerial war zone images.[67] Artificial intelligence and machine learning make decisions about what to attack, who to attack, when to attack.[68]

In 2025, the speed of cyber warfare has developed exponentially – we leave some decisions to machines because people can't decide quickly enough. Hence, military planners and critical infrastructures have incorporated AI into many operations. They excel at performing tasks, but they haven't included the ability to tell users why one decision is better than another, making some of their recommendations heretofore seem arbitrary or unreliable.

Universities have been contributing to the debate on AI in cyber warfare by developing artificial moral agents that can distinguish between good and bad and that can explain what they do.[69] Users can ask their smart information systems about why the systems accepted some recommendations and rejected others.[70] AI scientists, sponsored by governments and universities, have gone beyond developing explainability criteria[71] to developing a facility so that scientists can debate with AI systems the correct response to a cyberattack.

Applications in 2025

In their cyber war against the West, some foreign powers have been using artificial-intelligence-based applications to sow discord, spread disinformation, polarise society, attack critical infrastructure, including health systems, smart grids and nuclear power plants, and generally disrupt society, especially in NATO countries. Some foreign powers use AI in the automated reconnaissance of foreign networks, extraction of actionable intelligence, and subversion of adversaries' decision-making processes.

At least one foreign power uses AI systems to study the behaviour of social network users, and then designs and implements its own phishing bait. The artificial hacker is better at composing and

[67] U.S. Army Research Laboratory, "Artificial intelligence helps soldiers learn many times faster in combat", *Science Daily*, 27 Apr 2018.

[68] Dvorsky, George, "Hackers Have Already Started to Weaponize Artificial Intelligence", Gizmodo, 11 Sept 2017. https://gizmodo.com/hackers-have-already-started-to-weaponize-artificial-in-1797688425

[69] In 2014, the US Office of Naval Research offered a $7.5 million grant to a team from a number of universities – including Yale and Rensselaer – to develop robots with the capacity for moral reasoning. They intend to capture human moral reasoning as a set of algorithms, which will, quotes, allow robots to distinguish between right and wrong and to override rigid instructions when confronted with new situations. This appears to be a significant step towards so-called 'artificial moral agents' – intelligent systems endowed with moral reasoning that are able to interact with humans as partners.

[70] Raytheon BBN's Explainable Question Answering System will show users which data mattered most in the artificial intelligence decision-making process. Users can ask the system questions about chosen recommendations and discover why it rejected others.

[71] Explainability criteria are used to explain (briefly) the purpose of an algorithm, who developed it and who to contact for more information.

distributing more phishing tweets than humans, and with a substantially better conversion rate.[72]

The foreign power uses bots to hijack the public's perceptions of events and news. Bots, trolls and sockpuppets[73] can invent new 'facts' out of thin air leading to a polarised society and a culture of mistrust.[74]

A few companies have developed software to proactively detect and identify the bad guys from the moment they engage with client brand channels or protected accounts.[75] Under pressure from civil society organisations and parliamentary committees and in order to build trust with the public, those few companies have developed algorithms that explain why they took a particular action.[76]

In their own defence, some companies use AI to find and remove content from rogue governments, organised crime and terrorists from their websites and platforms. They use image-matching software to identify and prevent photos and videos from known terrorists from popping up on other accounts. They use machine-learning algorithms to look for patterns in terrorist propaganda. The big five companies (Amazon, Apple, Facebook, Google, Microsoft) have developed a shared database that documents the digital "fingerprints" of terrorist organisations.[77]

2. Vignette

Our vignette illustrates the ethical complexity of AI-powered cyber warfare in 2025. In that year, the Citizens Committee Against Nuclear Power (CCANP) has been campaigning against even a minimal proliferation of nuclear power plants because the energy they produce will be at least twice the cost of renewables over the 30-year-life expectancy of nuclear power plants. To sweeten the deal with the high cost and dangerous nuclear industry, the government has said it will bear the cost of disposing of the nuclear waste, which will, of course, be radioactive for thousands of years and threaten water supplies when the waste eventually burns through its glass and steel coffin in abandoned salt mines hundreds of metres below ground.

Do you think the use of a vignette helps to makes it easier for stakeholders to relate to the technology and its impacts?

The CCANP has said it would attempt to disable a power plant to show its opposition to the use of nuclear power. Meanwhile, the security services have detected rogue software embedded in the operating systems of many nuclear power plants, including that targeted by the

[72] Dvorsky, George, "Hackers Have Already Started to Weaponize Artificial Intelligence", Gizmodo, 11 Sept 2017. https://gizmodo.com/hackers-have-already-started-to-weaponize-artificial-in-1797688425
[73] Wikipedia defines a sockpuppet as an online identity used for purposes of deception.
[74] Singer and Brooking, *LikeWar*, op. cit., p. 208.
[75] SafeGuardCyber, p. 2.
[76] MacCarthy, Mark, "EU privacy law says companies need to explain the algorithms they use", CIO, 19 Oct 2017.
[77] Marr, Bernard, "Weaponizing Artificial Intelligence: The Scary Prospect Of AI-Enabled Terrorism", *Forbes*, 23 Apr 2018.

CCANP. Until recently, the software has been only monitoring the power plants' operations, but now has been causing sporadic stoppages.

Figure 7 Cyber attacks

Suddenly, one power plant shuts down. After weeks of claiming that it intended to take the power plant down, the CCANP claims it did not do it. Military and security strategists consider three options:

- The activists did it.
- A foreign power did it using the activists' threats as cover.
- Someone else did it and left fingerprints that would point to the foreign power.

The government wants to strike back, the military wants to strike back, the nuclear power plant operator wants to strike back, the majority of public opinion wants to strike back, but none is sure against whom they should strike. Soon after, a second nuclear plant is disabled. Critics say the government has not done enough to repel such attacks. Some critics say a counter-attack is now justified, but they are still none the wiser about who is behind the attacks.

Meanwhile, the leader-for-life of a foreign power has denied accusations in *The Guardian* that it was responsible. Instead, he claims it's all a Western plot to discredit the foreign power. Trolls and bots ensure the leader's denials ricochet around the Internet, overwhelming any rebuttals to the contrary.

3. Drivers

In 2025, many actors are engaged in cyberwar. Most are not in uniform. Governments, organised crime, terrorists and big companies engage in cyberattacks for a variety of reasons,[78] including the following:

Do you agree with these drivers? Are there any other significant drivers

[78] Singer, P.W., and Emerson T. Brooking, *LikeWar: The Weaponisation of Social Media*, Eamon Dolan/Houghton Mifflin Harcourt, New York, NY, 2018, p. 18.

Political drivers Nation states have been at real or de facto war. The foreign power mentioned above has been attacking many countries with the aim of disrupting them and, supposedly, strengthening its own power. In fact, the foreign power, with its armies of cyberwarriors, has become the most powerful nation on the planet, even though the economies of China and the US still dwarf that of the foreign power. *Cost savings in asymmetric warfare* Military budgets are constrained. Investing in AI-powered attack capabilities is less expensive than recruiting, training and maintaining fleets of aircraft and ships. If a few clicks on the keyboard can take down a power plant, is there any need for a bomber aircraft? With cyber weapons, armies don't need tanks, driverless or not, manned or not. A few hackers can cause millions of euro in damage. In AI-powered, asymmetric warfare, a few Davids can take down military Goliaths. AI is a game changer. *"Because we can"* The technological imperative is inexorable. Technology marches on. The tools for information warfare are widely available on the dark web.[79] Because they are, malefactors take advantage of them to wreak havoc, come what may. Meanwhile, big companies have invested in many different AI technologies and seek to sell their products no matter what consequences they produce. The technological environment evolves rapidly. *Trust and mistrust* AI-powered disinformation campaigns erode the public's trust in their governments. People do not know what is true and what is made-up. Malefactors use automated social engineering techniques to manipulate and divide populations against each other. The information society has become the disinformation society, with little accountability. *Modernisation of the military and intelligence agencies* The military, heretofore slow to recognise the change in state confrontation, is questioning its priorities and how to allocate its budget. Should it buy another aircraft carrier or use its budget to recruit hundreds of new cyber defenders and cyber warriors?	that should be included here?

[79] Not only are the tools widely available, they are often more advanced than what the good guys have. See Goodman, Marc, *Future Crime*, Corgi Books, 2015, p. 31: "criminals and virus writers are completely out-innovating and out-maneuvering the anti-virus industry". And at p. 55: "According to former FBI director Robert Mueller, there were at least 108 nations with dedicated cyber-attack units going after industrial secrets and critical infrastructure alike".

Fear

Fear of being overwhelmed by foreign powers, fear of defeat and fear of subjugation drive governments to invest in modern warfare. Fear of the unknown is a factor too. In the past, it was relatively easy to calculate how many aircraft or how many ships or tanks the enemy had using reconnaissance. Today, however, it is much harder to estimate how many cyber warriors the enemy has.

4. Barriers and inhibitors

Several barriers or inhibitors affect the pace of development of information warfare technologies in 2025.

Shortage of people with information warfare expertise

The big five have secured their grip on the world's economy by recruiting many of the world's data scientists. Government cybersecurity agencies are unable to match the salaries of the big five and struggle to find relevant expertise in information warfare.

Budgetary shortfalls

The cost of information warfare is high but Although the military does not need to struggle as much as the police to get their adequate share of the national budget, they do not get everything they need. Other national and international demands compete for a share of national budgets. Climate change is having a devastating impact not only on the environment, leading to droughts, failed agricultural yields, flooding of coastal cities, wildfires, earthquakes and super hurricanes, but also on national budgets are trying to contend with the ravaging impacts of climate change.

Black swan events

Black swan events – the x factor – constrain information warfare. The unauthorised release in January 2025 of e-mails between several large defence contractors revealed how they were stimulating warfare, conspiring to create crises to persuade politicians that governments should spend more on cyber defence. Not surprisingly, the public turned against the companies and demanded that politicians end their ties with the offenders.

Climate change

Humanity's destruction of the planet has begun to affect the extent to which countries are engaged in information warfare. By 2025, the ravages of climate change are felt everywhere, concentrating minds globally that more international co-operation (and fewer cyberattacks) is needed if civilisation is not to be completely undone.

Do you agree that these are likely to be the most significant barriers and inhibitors in 2025?
Are there any other barriers that should be mentioned?

5. Ethical, legal, social and economic impacts

Ethical impacts

Unintended consequences

Information warfare raises moral issues. The US and Israel developed Stuxnet specifically to target Iran's centrifuges, but an unintended consequence was the eventual release of the software into the wild, where it infected "thousands of computers across the world that had nothing to do with Iran or nuclear research".[80] Hence, critics in the US and Europe have questioned the development of cyber weapons, especially those that could cause collateral damage or have unintended consequences.

Do you agree that the ethical issues listed here are likely to be important in 2025?

Are there any other ethical issues we should include?

Some civil society organisations and leftist politicians call for a strategic and moral re-allocation of national priorities from combatting other countries and refocusing on the collective challenge facing civilisation from the ravages of climate change.

Employee pressures

Employees of the big five have pressured senior executives not to engage in the development of cyber weapons. Employee unions have successfully called upon senior management to install codes of ethics and codes of acceptable corporate practice.[81] The big companies are willing to install such codes as it helps them to forestall stricter regulatory oversight, while they know that ethical principles are sufficiently broadly written, they need not limit the company's ambitions, no matter what those ambitions might be. The codes enable the companies to hide behind their veils and pay lip service to corporate social responsibility and ethics.

Autonomous decision-making

For many researchers, giving machines the decision over who lives and dies crosses a moral line.[82] A key ethical issue remains: how much autonomy should AI solutions have? Informed opinion is divided: some say information warfare requires instant decision-making that obviates the possibility of human intervention. Others say that some untoward events involving AI (e.g., driverless cars causing fatal accidents, robots turning on their 'masters' or malfunctioning drones blowing up school buses) show that human intervention must always be possible. In any

[80] Singer, P.W., and Allan Friedman, *Cybersecurity and Cyber War: what everyone needs to know*, Oxford University Press, 2014.

[81] Harwell, Drew, "Google bans development of artificial intelligence used in weaponry", *The Washington Post*, 7 June 2018.

[82] Sample, Ian, "Thousands of leading AI researchers sign pledge against killer robots", The Guardian, 18 Jul 2018.

event, there is widespread agreement among stakeholders and the public on the explainability principle (algorithms must be able to explain what they are doing and whom to contact for more information), even if the principle is difficult to implement.

When to retaliate and what is a proportionate response?

For several years, there has been much debate about when to retaliate against cyberattacks and who should do so. The US and European governments have warned companies and citizens not to take the law into their own hands. They should share any information about attacks they've suffered with others in their sector and, especially, with national cybersecurity agencies, but this policy has not been an adequate response, in part, because there are so many cyberattacks and because national cybersecurity agencies are unable to defend companies and citizens against all such attacks. Hence, companies and governments have, therefore, adopted a different policy, i.e., it is acceptable to retaliate in certain circumstances. Government officials and companies have set up working groups to debate under which circumstances and how measured retaliatory responses should be against different types of attacks. How should we act when we have only 75 per cent certainty of who is likely responsible for a cyberattack?

A dangerous space

Information warfare involves virtually everyone using the Internet, either as a victim or a warrior. The days of the uninvolved, unattached surfer have long gone. The Internet has become a dangerous space that you enter at your own risk. Decision-makers, from parents to parliamentarians, are confronted with ethical dilemmas every day. Should children and vulnerable people be advised to limit their use of the Internet to the absolute essentials? Should they be trained to recognise aggression and how to respond? How do we spot manipulation? Should we embed algorithms with morality – i.e., to do good and to shun evil – when questions inevitably arise about what is good and evil.

Loss of the high road, loss of trust

Foreign powers say that those in the US and Europe who cast aspersions about the conduct of information warfare by foreign powers are hypocrites, as the US and its NATO allies have been caught out deploying AI-powered malware, just like them.

With so many countries engaged in information warfare to a greater or lesser extent, trust between countries has been a casualty. Foreign powers may deny they are responsible, but the evidence shows otherwise.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Legal impacts

The prevalence of artificial intelligence in information warfare raises many legal issues and has many impacts in 2025.

Definition of warfare

The definition and scope of warfare has been the subject of much debate in the US and Europe. If an aircraft from a foreign power bombed a nuclear power plant in Connecticut, the US would rightly view such action as an act of war and retaliate accordingly. However, if some foreign power's malware disabled the plant, the reaction of the US might not be so clear. [83]

The European Commission has been reluctant to fund military research in the use of AI in its recently concluded Horizon Europe (HE) research. It has been relatively easy to proscribe use of the HE research budget to fund the advancement of killer robots, killer drones, autonomous submarines. However, it has been harder to draw red lines against cyber weapons – the distinction between defensive tools and offensive weapons has blurred given the potential of many tools and information systems for dual use.

Figure 8 Cyber weapons

Liability

With so many players engaged in warfare, ascribing liability continues to pose legal challenges. Some argue that policymakers who turn a blind eye should be held as liable as the companies that develop algorithms that power bots, denial-of-service attacks, ransomware and

Do you think these will be the key legal issues in 2025?

Are there any other legal issues that we should include?

[83] UK government minister Jeremy Wright has been quoted as saying: "If a hostile state interferes with the operation of one of our nuclear reactors, resulting in widespread loss of life, the fact that the act is carried out by way of a cyber operation does not prevent it from being viewed as an unlawful use of force or an armed attack against us", Sky News, 23 May 2018.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

other malware. Flawed policies lead inexorably to an amplification of warfare. Others blame politicians, the military-industrial complex, right-wing ideologues, and the media for fuelling fears. Still others ask, who should be liable when AI acts on its own? The programmer? The data scientist? The copyright or patent owner? The supplier of the technology? The service provider? Or perhaps the owner of the dataset(s) on which the algorithm was trained? Will companies pursue certain technologies if they are held liable for their misuse?

The multiplicity of players in the AI chain dilutes accountability. Although there are no formal declarations of war, governments, big companies and rogue actors are engaged in information warfare with consequences every bit as deadly as if a foreign aircraft flew over the proverbial nuclear power plant in Connecticut and blew it to smithereens.

New legislation, new regulation

The EU's General Data Protection Regulation (GDPR) and Police Directive have generally proved remarkably fit for purpose and address most aspects of data, which fuels AI. However, AI raises more than data protection issues. It raises a range of ethical, social, political, economic and other issues too. Data protection authorities have engaged in some mission creep, expanding their remit from regulating pure data protection issues to addressing ethical issues too.[84] Even so, some European regulators have recognised that AI requires special legislation and regulation.[85] On the recommendation of the European Commission, the EU Council and Parliament created a new European Regulatory Agency for AI (ERAAI) in 2024, following six years of intense debate about the agency's remit and purview. In the end, the vast power of the big five convinced legislators of the need to exercise some political control over their power, as exercised through their algorithms.

Rules of information warfare

In 2004, the UN set up the Group of Governmental Experts on Information Security to agree voluntary rules for how states should behave in cyberspace. Its fifth meeting, in 2017, ended in a stand-off. The group could not reach consensus on whether international humanitarian law and existing laws on self-defence and state responsibility should apply in cyberspace.[86] The stand-off continues in 2025.

One country attempted to promote an international treaty on the rules of engagement in cyber warfare. The power grid and water

[84] 40th International Conference of Data Protection and Privacy Commissioners, Declaration on Ethics and Data Protection in Artificial Intelligence, Brussels, 23 Oct 2018.

[85] Nemitz, Paul, "Constitutional democracy and technology in the age of artificial intelligence", *Philosophical Transactions of the Royal Society*, Oct 2018.

[86] Taddeo, Mariarosaria and Luciano Floridi, "Regulate artificial intelligence to avert cyber arms race", *Nature*, Vol. 556, 19 Apr 2018, pp. 296-298

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

infrastructure should be off-limits to any attacks.[87] Despite the favourable coverage in much of the world's media, few countries were willing to subscribe to a treaty that limited their powers. In any event, when the government mooted such a treaty in 2020, all of the major cyber powers had already embedded malware in their enemies' power grids.

The UK and some other countries have declared that they view the use of cyber technologies to interfere in another state's elections as contravening international law and norms; consequently, affected states should take whatever action they see fit.[88]

The legal limits of solidarity

Article 5 of the Washington Treaty, which established NATO in 1949, states that "The parties agree that an armed attack against one or more of them in Europe or North America shall be considered an attack against them all." The limits of solidarity were amply illuminated when Russia launched a denial-of-service attack against Estonia in April 2007, hitting banks, media web pages, the government website. Estonia called upon NATO for assistance, but the other members didn't think Article 5 applied.[89]

Foreign powers and other malefactors have been successful in exploiting the general perception that cyber war is somehow different from conventional war even though the consequences may be the same or, in many cases, much worse, spilling outside defined battlefields and traditional war zones.

Governments have been cautious about attributing attacks, in part because their origin can be hard to trace, as depicted in our vignette, and in part because they have not wanted to reveal how they have tracked or penetrated the groups. But the US, UK, Canada, Australia, France and other countries changed their tactics several years ago and began naming names and the countries of the perpetrators of cybercrimes.[90]

Social impacts

The threat of counterstrike requires knowing who launched the initial attack, a difficult thing to prove in cyberspace, especially in a fast-

Do you think these are likely to be important social impacts created

[87] The 2016 EU directive on 'Security of Network and Information Systems' provides criteria for identifying crucial national infrastructures, such as health systems or key energy and water supplies that should be protected. The same criteria could be used to define illegitimate targets of state-sponsored cyberattacks.

[88] Martin, Alexander J, "UK begins to formalise its legal approach to cyber war", Sky News, 23 May 2018. https://news.sky.com/story/uk-begins-to-formalise-its-legal-approach-to-cyber-war-11382545

[89] Singer, P.W., and Allan Friedman, *Cybersecurity and Cyber War: what everyone needs to know*, Oxford University Press, 2014, pp. 348-349 (of iPhone 6 version).

[90] Nakashima, Ellen, Michael Birnbaum and William Booth, "U.S. and its allies target Russian cyber spies with indictments, public shaming", *The Washington Post*, 4 Oct 2018

moving crisis, as in our vignette.[91] Deterrence does not work in all circumstances, e.g., where non-state actors are major players in cyberspace. Not all are rational or predictable actors. Many voters regard a foreign power's flagrant manipulation of elections as an act of war. Warfare is not just about blowing up bridges and railway lines anymore; it is also about discrediting politicians, planting vast amounts of misinformation, so that voters and the public are unable to distinguish truth from lies. A lie repeated hundreds of times is more powerful than a fact-checker repeated once. One of the main defences in a state of information war is surveillance. We should expect surveillance to increase, but by 2025, there has already been so much surveillance, that most people are not concerned by more. A decade ago there was serious opposition to national biometric databases with records of everyone's DNA, fingerprints, photo identity. Now, not so much. Social cohesion has been a major casualty of information warfare. People don't know whom to trust or what to trust, even if they are aware of the political struggles underlying information warfare. Some would argue that individual autonomy has been another casualty. If citizens' voting intentions can be swayed by information warfare, autonomy is so much roadkill.	by these technologies in 2025? Are there any others you think we ought to include?
Economic impacts The cost of recruiting a cyberattacker is relatively low compared to the cost to organisations in defending themselves against attacks. The value of the AI-powered cybersecurity applications has soared from $1 billion in 2016 to more than €25 billion in 2025.[92] In other words, there is a huge asymmetry in the cost of attacks versus the cost of defence. Despite the huge expenditures, the UK Parliament's Public Accounts Committee and the US Government Accountability Office (GAO) have revealed that nearly all of the allies' weapons systems have cybersecurity vulnerabilities.[93] Cybersecurity represents a major cost to all organisations. On the other hand, the cybersecurity industry is a correspondingly big employer, with a high-tech workforce for whom there is a big demand no matter how high salaries are. The soldiers in the information wars of 2025 seem like light-years away from the raw soldiers who fought in the trenches of World War 1. 2025's soldiers use their minds more aggressively, and they come at a price.	Do you agree with these economic impacts? Are there any other security and economic impacts that you think will be particularly important in 2025?

[91] Singer, p. 412.

[92] Taddeo, Mariarosaria and Luciano Floridi, "Regulate artificial intelligence to avert cyber arms race", *Nature*, Vol. 556, 19 Apr 2018, pp. 296-298

[93] Zachary Fryer-Biggs, Center for Public Integrity, "'Nearly All' of the Pentagon's New Weapons Systems Are Vulnerable to Hacking", *The Daily Beast*, 10 Oct 2018.

Figure 9 Cybersecurity

Information warfare encompasses not only nation-states but also big companies attacking their rivals whether they are in the US, Europe, China or anywhere else.[94] Artificial intelligence has made cyberattacks such as identity theft, denial-of-service and password cracking more powerful and more efficient. AI systems can steal money, cause emotional harm and kill people. They can deny power supply to hundreds of thousands of people, shut down hospitals and compromise national security.[95]

AI helps states and their attackers customise attacks. AI systems help gather, organise and process large databases to connect data points, making attacks easier and faster to carry out. That reduced workload may drive perpetrators to launch lots of smaller attacks that go unnoticed for a long period of time – if detected at all – due to their more limited impact. AI systems draw information together from multiple sources to identify people who are particularly vulnerable to attack. [96]

6. Recommendations for a desired future and avoiding an undesired future

In this section, taking into account our scenario, we present recommendations to EU and MS policymakers to help us (as a society) to reach the future we want in 2025 and to avoid the future we don't want.

Do you agree with these recommendations?

[94] Sheera Frenkel, Nicholas Confessore, Cecilia Kang, Matthew Rosenberg and Jack Nicas, "Delay, Deny and Deflect: How Facebook's Leaders Fought Through Crisis", *The New York Times*, 14 Nov 2018. https://www.nytimes.com/2018/11/14/technology/facebook-data-russia-election-racism.html

[95] Straub, Jeremy, "Artificial intelligence cyber attacks are coming – but what does that mean?", *The Conversation UK*, 28 Aug 2017. http://theconversation.com/artificial-intelligence-cyber-attacks-are-coming-but-what-does-that-mean-82035

[96] Ibid.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Other countries in the EU should emulate the actions of Estonia and Sweden to create "whole-of-nation" efforts intended to inoculate their societies against viral misinformation, including citizen education programmes, public tracking and notices of foreign disinformation campaigns and enhanced transparency of political campaign activities,[97] so that citizens are informed about efforts to undermine their democracies. The European Commission should recognise that cyberattacks are a form of warfare – information warfare, but no less warfare for that. The EC should define cyber warfare. Its definition should include attacks by nation states, crime gangs, terrorists against critical infrastructures and its impact on society and major social groups. Governments should reveal the full extent of cyberattacks, where they can be traced, [98] but it is not sufficient to merely expose a rogue state's conduct; law enforcement authority should seek to arrest those who broke the law.[99] Some retaliatory action is needed. For example, in the vignette, in retaliation to the shut-down of the two nuclear power plants in the UK in 2025, the US and UK could demonstrate their ability to turn off the power in the foreign power's capital city with a one-minute black-out. They could threaten a longer black-out if the foreign power continues to attack their nuclear power plants.[100] But other forms of retaliation are possible too, e.g., exposing the wealth of the foreign power's leader hidden in the vaults of Zurich, the Cayman Islands and other such havens.[101] Exposing what the leader-for-life and his cronies do for entertainment is another form of retaliation. The EC should provide funding for studies on information warfare via the European Defence Fund and the forthcoming Horizon Europe research programme and, in particular, how AI is being used to spread misinformation, hate crimes and lies, especially to undermine elections, and how to assess and what to do about the resulting social impacts and what the EU should do about such activity.	Are there any others that you think we should include?

[97] Singer, PW, and Emerson Brooking, "The election hackers are back – and they're starting with the US midterms", *The Guardian*, 26 Oct 2018.l

[98] Wintour, Patrick, "UK accuses Kremlin of ordering series of 'reckless' cyber-attacks", *The Guardian*, 4 Oct 2018.

[99] Nakashima, Ellen, Michael Birnbaum and William Booth, "U.S. and its allies target Russian cyber spies with indictments, public shaming", *The Washington Post*, 4 Oct 2018.

[100] Ardehali, Rod, "Britain 'rehearses cyber-strike to black out Moscow' in the event of Russia attacking the West as thousands of UK troops stage biggest war-games exercise in a decade", *The Daily Mail*, 7 October 2018. https://www.dailymail.co.uk/news/article-6248427/Britain-rehearses-cyber-strike-black-Moscow.html. In fact, the US has recently started to take pre-emptive actions. It temporarily cut off Internet access by the Internet Research Agency, the St Petersburg-based troll factory implicated in disrupting the US 2016 presidential election. Nakashima, Ellen, "U.S. Cyber Command operation disrupted Internet access of Russian troll factory on day of 2018 midterms", *The Washington Post*, 26 Feb 2019.

[101] Marks, Joseph, "U.S. should counter Russia and China hacking with its own influence operations, think tank says", *The Washington Post*, 1 Feb 2019.

European policymakers should not be in reactive mode to the impacts of AI in information warfare. They should be pro-active, considering a wide range of measures, including offensive measures against individual attackers sponsored by governments as well as the governments themselves. The EC, ENISA, national cybersecurity agencies and industry should develop a co-ordinated strategy for countering attacks against individual companies and to what extent they can engage in retaliatory activities. Big companies, such as Amazon, Apple, Facebook, Google and Microsoft, are more capable than most countries in taking more aggressive action against entities abusing the Internet and engaged in misinformation campaigns and cyberattacks, such as the attack on the nuclear power plants in the UK depicted in the vignette. Governments alone do not have the resources to counter all attacks, but there should be a consensus in the EU and elsewhere in what instances companies can engage in offensive strategies. Compared with traditional armed conflict, the rules of information warfare are not well-defined. The European Commission and/or the United Nations should develop such rules, especially applicable to the private sector. We need the information warfare equivalent of the Budapest Cybercrime Convention[102]. Tech firms need to step up investment in content moderation; take down those engaged in harassment and foreign influence operations; test their products for dual-use capabilities before they are deployed, not just for cybersecurity vulnerabilities, but misuse by attackers; label bots in order that humans can tell when they are interacting with a machine online; and implement measures to foil the next generation of AI used in sophisticated chatbots and faked imagery.[103] Politicians and diplomats should call for an end to information warfare, so that more resources can be channeled to combatting climate change.	

[102] https://www.coe.int/en/web/cybercrime/the-budapest-convention

[103] Singer, PW, and Emerson Brooking, "The election hackers are back – and they're starting with the US midterms", *The Guardian*, 26 Oct 2018.

5. Third scenario: Predictive policing in 2025

Figure 10 Third scenario

1. Introduction

This scenario focuses on AI-powered predictive policing in the year 2025. It considers the issues such applications raise, their impacts and how policymakers especially should respond to those prospective impacts.

The structure of this scenario differs somewhat from the others in that it starts with a vignette which is then dissected in the sections that follow. So it starts…

2. Vignette

In 2025, many police forces across Europe are adopting predictive policing technologies in response to cuts in human resource budgets. Such cuts inevitably led to a rise in crime rates. Many law enforcement authorities (LEAs) began experimenting with different predictive policing technologies as a way of cutting crime before it happens. After some false starts, such technologies have evolved as remarkably as facial recognition technologies. Smart information systems, notably artificial intelligence algorithms, are within the reach of all European LEAs, who now can feed such systems with the vast swathes of data to which they have access. In a manner that is both intelligent and provides usable information in real time, LEAs have been experimenting with different applications. Some of these have been developed in-house by the national forces, some have been developed through the European Commission's Horizon Europe research programme, but many are the result of collaborations with private sector players. In some cases, these private initiatives include or result in proprietary data of benefit to the private sector partners.

As one would expect, some approaches and technologies for predictive policing have proven to be better than others. The intelligence-led policing approaches trialled by Pol-Intel in Denmark[104] have served as models of police access to and use of many disparate data sets. The more ambitious applications go beyond accessing data to using those data to make predictions regarding incidents of future crime. Most predictive policing applications have drawn on location-based data to define increasingly localised "hot spots" on which the police should focus attention at particular times, while others draw on personal data to identify likely and repeat offenders[105]. Other applications aim to predict likely and repeat victims of crime in cases such as domestic violence, or those at risk of becoming offenders in the future. Still other predictive policing applications have turned their attention from visible street crime to the less visible white-collar crimes, including money-laundering, tax evasion, fraud and cybercrime. Some researchers are using these technologies to draw together demographic, census and other social data to determine what factors are most likely to induce someone to commit crime. The answers to such questions are expected to make possible early, large-scale interventions where communities and/or individuals are at risk.

Predictive policing applications must have measurable success factors. Typically, this is a matter of rising or falling reports of crime, but this is an unstable metric. At its heart is a mere correlation, which doesn't prove a causal link between the application and the number of reports. Hence, a decline in reported crime might have come about through using the application, but it might equally be a result of demographic changes. It is possible that reliance on the application has reduced the efficacy of police responses such that many no longer bother reporting crimes as they know that the reports won't be acted upon. Equally, some applications have been reported as helping the police determine which crimes are worth a response. In some areas, thanks to local press reporting, it is widely known that burglaries will usually not merit a police response, and so actual burglaries have increased in number while the number of reported burglaries has declined. On the other hand, the applications may be so successful that police are effectively anticipating crimes and arriving in time to deter the potential criminal from carrying out his or her plans. This is plausible given efforts to streamline the online reporting process, itself aided by data analytics and AI allowing for a smooth and fast process for victims and other to report crimes.

[104] Bjørnholdt, K, "Ny it skal hjælpe politiet med at fange forbrydere", Dansk Politi, Copenhagen, 22/12/2016. https://www.dansk-politi.dk/artikler/2016/december/ny-it-skal-hjaelpe-politiet-med-at-fange-forbrydere. Accessed 04/01/2019. Kulager, F, "Vi prøvede dansk politis nye, kontroversielle datavåben. Det tog os ti minutter at opklare en sag om hashsmugling", Zetland, Copenhagen, 20/06/2018. https://www.zetland.dk/historie/sop1JZkz-aOZj67pz-e4a92. Accessed 04/01/2019.

[105] Norwegian Board of Technology, "Predictive Policing: Can data analysis help the police to be in the right place at the right time?", Norwegian Board of Technology, Oslo, 2015.

While some of the public feared a move to "Minority Report" policing, in which a computer informs police who is about to commit a crime and then that person is arrested moments before the act, this has not happened. Indeed, the police are adamant that any computer prediction regarding likely crime hot spots or offenders is fed as information to a team of analysts who then combine that data with other information before advising patrols. This prevents policing by algorithm from becoming the norm. However, cuts in police funding have reduced the number of available analysts, and the remaining analysts have been noticing that the number of false positives (indications that a crime will occur in an area where no crime takes place) is falling with each year and worry for the future of their jobs. In 2020, there was only one information analyst working for the LA Police Department. Furthermore, budget cuts have pushed many officers with good local knowledge into early retirement. New officers, lacking this knowledge, are content to rely upon the predictive policing system. This has led to fears of automation bias in which officers trust the system despite evidence to the contrary[106], and despite the training, introduced in 2020, to rectify this.[107] Nonetheless, there remains a tension as to how best to act when the system recommends one course of action and the officer disagrees with this recommendation, leading to some complaining that they are being treated as robots.

International comparisons do not end with the numbers of analysts. Many cities in the United States have been aggressive in pursuing predictive policing, particularly after funding was increased shortly after Trump was re-elected in November 2020. Incarcerations have increased, but there is no sign of a change in the demographic composition of the prison population, which is overwhelmingly African-American. China has also been aggressive in developing predictive technologies following the widespread integration of the Social Credit System which incorporates all data on a person, including bank records, medical records and educational attainments[108]. Facial recognition on CCTV is now standard in most Chinese cities, although there is insufficient recognition by the Chinese authorities of the problem of false positives. The general approach is one of "better safe than sorry", leading again to a suspected (albeit unreported) rise in the prison population. Owing to Chinese information sharing protocols, it is also not certain what the ethnic composition of that population looks like, but there are reports that some communities such as the Uighurs have been all but decimated in recent years as they are arrested on the basis of

[106] Cummings, M.L., "Automation bias in intelligent time critical decision support systems", in: AIAA 1st Intelligent Systems Technical Conference, 2004, pp. 557–562.; Wickens, C.D., Clegg, B.A., Vieane, A.Z., Sebok, A.L., "Complacency and Automation Bias in the Use of Imperfect Automation", Human Factors, Vol. 57, 2015, pp. 728–739. https://doi.org/10.1177/0018720815581940.

[107] Goddard, K., Roudsari, A., Wyatt, J.C., "Automation bias: a systematic review of frequency, effect mediators, and mitigators", Journal of the American Medical Information Association, Vol 19, 2012, pp. 121–127. https://doi.org/10.1136/amiajnl-2011-000089, 2012.

[108] Creemers, R., "China's Social Credit System: An Evolving Practice of Control", SSRN Scholarly Paper No. ID 3175792, Social Science Research Network, Rochester, NY, 2018.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

a likelihood of committing a crime[109]. Finally, efforts at introducing predictive policing in some South American cities, such as Bogota in Colombia, have exacerbated perceived biases as the focus remains on preventing crimes against the wealthy, while police ignore victims from less affluent areas.

Figure 11 CCTV - Image Credits: Niv Singer, Flickr (CC BY-SA 2.0)

Europe has been slower than China and the US in adopting predictive policing technologies, partly owing to the human rights frameworks such as the European Charter for Fundamental Rights and the European Convention on Human Rights, both of which are reinforced by laws such as the General Data Protection Regulation, which is still seen as an effective means of regulating the use of personal data across society. The so-called EU Police Directive, however, gives LEAs more flexibility in processing personal data. This regulatory framework combined with the Horizon Europe research programme, begun in 2020, boosted funding for counter-terrorism efforts and cybersecurity.

While there has been investment in police use of these technologies, criminals have not been idle. LEA cyber sleuths have uncovered applications used by criminal gangs to predict where the police will be at any time of the day or night, often drawing on the same data sets used by the police, made public in the name of transparency and democratic accountability. Others have been found on the dark net offering significant sums to hackers who can reverse-engineer police systems to indicate which parameters are used to predict crimes in order that they can better avoid detection.

The police find themselves caught between a rock and a hard place. The press is critical of any reports of rising crime and cynical of reports to the contrary. The police do not need to be reminded of their duty to do all that

[109] Rollet, C., " In China's Far West, Companies Cash in on Surveillance Program That Targets Muslims", *Foreign Policy*, 20128. Samuel, S., "China Is Going to Outrageous Lengths to Surveil Its Own Citizens", *The Atlantic*, 2018. https://www.theatlantic.com/international/archive/2018/08/china-surveillance-technology-muslims/567443/. Samuel, S., "China Is Treating Islam Like a Mental Illness", *The Atlantic*, 2018. https://www.theatlantic.com/international/archive/2018/08/china-pathologizing-uighur-muslims-mental-illness/568525/

is reasonable to prevent crime, but the debate within society as to what is reasonable, including which databases can be routinely accessed, rages on with an apparently fickle public opinion swinging wildly in polls depending on the latest scandal.

3. Drivers

Various drivers have impelled the development of technologies used in predictive policing in 2025, among which are:

Do you agree with these drivers? Are there any other significant drivers that should be included here?

Resources

Ever tighter squeezes on funding have led to a decline in the number of officers over the past decade while investment in technology has increased. AI is often treated by politicians as a panacea to limited public funds. There is some dissension in the ranks, as many officers can see that while the police budgets are shrinking, the technology firms developing AI applications seem to be thriving. If police budgets for human resources have been declining, the quantity and quality of data processed by the police has not. In fact, there is now so much data available from so many different sources that the police would be overwhelmed by it all were it not for artificial intelligence.

Public perception

Given the increased data available, there is a concern that the police miss intervening in cases where they had the relevant information in advance but did not process it in time. This is widely seen as a dereliction of duty that no Chief Constable wants to see on her watch. The public view of the police is ambivalent at best and there is a high level of expectation on the police and their use of technology. After all, if a member of the public can prove that his phone is in his neighbour's house through using tracking apps like Prey, he wonders what is there to stop the police from entering the home and retrieving the phone? His reasoning leads him to conclude that the police are either unwilling to help him or that they are hopelessly out of date.

International and technological factors

As noted above, Europe has been less aggressive in employing predictive technologies than other countries, notably the US and China, which have considerable resources and public support to invest in these technologies. Many European data scientists have already migrated to one of these countries to work on systems that receive minimal attention from European politicians These data scientists opine that we must follow where technology leads. This divestment of talent, coupled with the mixed results of the Horizon Europe research projects, has led some European police forces to buy technologies from US and Chinese companies, although they are uncomfortable with the fact that these were likely developed in a manner not consistent with European law. Furthermore, there is the ever-

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

present fear that US or Chinese intelligence agencies will infiltrate these systems through backdoors to spy on their European counterparts.

The last few years have seen a remarkable proliferation of AI ethical frameworks sprouting up everywhere. While these may not actually improve practice – because they are naïve, weak, compatible with authoritarian practice, or just used as fig-leaves -- nevertheless they serve as a driver because some police forces are investing in data scientists, while others are developing their own predictive technologies in-house.

4. Barriers and inhibitors

While there have been several drivers pushing the development of predictive policing technologies towards their current state in 2025, this development has not always been straightforward. There have been hurdles that have impeded progress. These have included:

Social factors

Media coverage of increasing use of technology was rarely positive and, while the intended target was often politicians, it was the police who suffered from the adverse coverage. In particular, the press noted the lack of change in the demographics of those arrested and imprisoned. While some have argued that a turn to computerisation in detecting and predicting crime would lead to greater objectivity, this appears not to have been the case.

Even where the predictive capacities of the applications have been more effective, these were met by the equal capacities of criminals who were able to emulate the predictive tools and hack into them directly. This has become part of the continuing escalation of methods used by the police and criminals to stay one step ahead of each other. Most applications are in a constant phase of beta-testing as by the time they are sufficiently stable to be rolled out on a wide basis their method has been cracked and they are no longer as effective.

There has been some marked resistance to change from within the police forces themselves. Some of this has been resolved through generational change as the post-millennial generation who grew up on smart phones have come of age and started to enter the workplace, but some resistance remains.

Other factors have been disrupting LEAs. Some LEAs have lost a quarter of their staff through retirement in the last five years. Such big losses have prompted senior officers to consider more carefully the work force they want for the technological challenges of the 21st century.

Economic factors

Do you agree that these are likely to be the most significant barriers and inhibitors in 2025?
Are there any other barriers that should be mentioned?

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Resources have been a driving factor in the development of predictive applications but, paradoxically, they have also held back some aspects of development. There has been a chronic shortage of computer scientists developing tools, and a shortage of analysts with the abilities to effectively use those tools. This is largely due to the inability of the public services to compete with private organisations, especially those working in similar areas of technology in other countries. Limited funding has also led to less reliable datasets and tools than would be ideal, with the result that their accuracy and efficiency sometimes leaves a lot to be desired. Despite this, for some, an 80% conviction rate is good enough, and many are becoming increasingly over-reliant on the systems that have led to a positive (although not a virtuous) feedback loop.

Even if they are convinced of the efficacy of AI supported predictive policing, a major inhibitor for LEAs is finding data analysts and scientists. The big five are scooping much of the available talent. Some LEAs are trying to overcome professional shortages by partnering with universities and taking PhD students as interns. The EU and Member States are well aware of the shortages of talent and, as a consequence, some MS have established national AI programmes aimed at cultivating data analysts and scientists.

Political factors

The lack of funding is due to continued attempts to rein in public spending in the post-2008 world. Some politicians worry about the press drubbing them and the police for arresting people for crimes they haven't committed yet. Some sceptics criticise the lack of effective and convincing metrics demonstrating the success of the technologies.

Legal and regulatory factors

To ensure police accountability in the use of data analytics and their big databases, parliament adopted laws and regulations that, among other things, made explainability the default mode for algorithms. Politicians had to balance concerns about individual privacy and data protection with the efficacy of police operations. The police were concerned that excessive transparency would give criminals better insight into police methods and, as it turned out, police concerns were justified. Consequently, a committee of the European Parliament has been investigating and debating whether algorithms developed for or used by LEAs should be compelled to have the same standards as others if organised crime benefits from the tiniest scrap of information.

One solution to the stricter regulations imposed by Brussels and national governments on artificial intelligence has been the outsourcing of some technologies to private companies. Without incentives, these companies only complied with the minimum requirements of the law, to the chagrin of many LEAs who knew these companies should be doing more to help them in the fight against organised crime. The press saw this outsourcing as having the effect of blurring the borders between policing and the corporate world even more than was already the case in the early 21st century.

5. Ethical, legal, social and economic impacts

In 2025, the benefits of predictive policing technologies are starting to be felt, even though there is still considerable public discussion as to whether these are strictly attributable to the technologies or other factors. Nonetheless, their use has been part of a marked shift in society as noted below:

Ethical impacts

Older police officers resent the tighter constraints on their actions compared to when they started their careers. They feel the so-called "smart" information systems that tell them where to go and what to do, are undermining their own skills, experience and talents in responding to crime. Older policemen don't seem to recognise how organised crime has shifted away from street crime to more high value crime in money-laundering and cybercrime. At the same time, there is clearly greater accountability and transparency in policing as bodycams record every move of every officer and individual officers are frequently held to account over why they did or did not intervene in a particular situation.

Do you agree that the ethical issues listed here are likely to be important in 2025?
Are there any other ethical issues we should include?

Figure 12 Police investigations

Civil society organisations protest that predictive policing technologies are an affront to Europeans' fundamental rights. There is much debate within police ranks and others about whether when a police officer responds to an algorithm that has 80% predictive capabilities, she is infringing on a person's civil rights by treating him as a suspect on the basis of a statistical calculation rather than his *doing* anything to warrant suspicion. At the same time, if she fails to act on the prediction, is she thereby failing to

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

uphold the civil rights of potential victims? She has no misgivings: her system justifies her suspicions because the suspect has committed crimes previously.

This fallaciously assumes that statistical calculations don't apply to things people have done. Whether or not something warrants suspicion depends on how highly correlated/causally correlated it is with the commission of a specific crime and, at the same time, how little it is correlated with innocuous behaviour. Most if not all current predictive techniques rely very heavily on crime and police data (e.g. arrests etc) which are about suspicious things people have done. Big data in policing is still in its infancy. One upshot of the current reliance on police data is that those with a profile in a police database are much more likely (even, the only ones) to be identified as a future threat. This creates a ratchet effect for those in the system. It also means predictive techniques are not able to detect first-time offenders. This makes people with no record easy targets for exploitation by criminals. It is also bad for domestic abuse homicide victims, whose perpetrators often have no record.

More positively, prior to the implementation of predictive technologies, individuals were already being stopped and searched, and arrested, sometimes for spurious reasons. The aforementioned increase in accountability has shed light on discriminatory stop-and-search practices. Overall, predictive policing technologies have reduced some discriminatory practices and embedded others, such as an algorithm that focuses more on street crime than corporate malfeasance.

The public discussion that accompanied the widespread introduction of these technologies helped ensure that the explainability regulations in Europe were fair, ethical and sensitive to privacy concerns. Public pressures led to the establishment of independent oversight bodies in the Member States to monitor police use of smart information systems.

While media attention has focused on the police use of predictive applications, some members of the fourth estate have focused on corporate responsibility. Since social media giants collect reams of data, they are frequently able to identify child sex offenders or people involved in domestic abuse. However, this information is rarely turned over to the police. Questions are being asked in national legislatures about the social responsibility of these organisations.

Ethical issues have risen high on policy agendas within LEAs themselves as well as in their oversight bodies. LEAs recognise that to improve trust with the public, they need to be more transparent about their priorities and how they operate. Similarly, progressive LEAs expect the AI systems they use to be explainable and not simply black boxes. In other words, the AI systems used by LEAs should be capable of interrogation, should explain their purpose and whom to contact for more information.

Legal impacts

A key problem with the development of legal and regulatory frameworks in keeping up with technological development is that policy and lawmakers often do not understand the technologies. Technological development is happening faster than the passage of laws and has been impeded by the time lawmakers need to understand recent developments and the subsequent legislative process. The GDPR, which came into effect in 2018, remains generally fit for purpose regarding personal data, but with the aggregation of databases, it is increasingly rare to find data that cannot in some context or manner be used to identify a living person. The most applicable legislation for LEAs remains the Police Directive, which has meant that LEAs did not need to seek informed consent when they were investigating persons of interest. With so many AI-powered applications available online, prohibitions against automated decision-making affecting the rights of data subjects have become impossible to enforce except in a few high-profile cases like those against Google and Facebook in 2020-21. That so many enterprises see that it is impossible to enforce some provisions of the GDPR has had the predictable consequence of diminishing trust in the law even from law-abiding companies and citizens.	Do you think these will be the key legal issues in 2025? Are there any other legal issues that we should include?
Social impacts Criminals seek advantage over LEAs by exploiting new technologies before the police are able to put counter-measures in place. The nature of crime is changing. The police have been shifting their focus from street crime, which is particularly subject to some of the blunter forms of predictive policing technology, to organised crime and white collar crimes, including money-laundering, fraud, online scams and hacking. While organised crime gangs are aware of predictive policing technologies (which receive a lot of attention in the newspapers), the public generally has a low understanding of such technologies and their possible negative impacts. The public is bombarded with so much information (and disinformation) about new technologies that the public has become jaded. The powers of new technologies have ceased to spark wonder. The majority of the public accept these measures as just part of the cost of living. The public has already learned to cope with the substantial levels of surveillance in society – on the streets and in cyberspace. Some people claim that they have altered their behaviour, to appear as conformist as possible, as these days, they do not know what will land them in some police database. Better to play it safe.	Do you think these are likely to be important social impacts created by these technologies in 2025? Are there any others you think we ought to include?
Economic impacts We have already noted the savage cuts in police budgets, also of note is the shift in budgetary priorities from police officers to more data analysts. As the number of officers falls, so the reliance on AI grows, and as the reliance on AI grows, so the same work (or at least similar) is apparently achieved with fewer officers, and so funding declines further. One solution has been to outsource certain tasks, such as facial recognition, to the	Do you agree with these economic impacts? Are there any other economic impacts that you think will be particularly important in 2025?

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

private sector, as the US has done for several years. However, outsourcing has largely been discredited.

6. Mitigating the negative and acting on the positive impacts

For some people, predictive policing was an easy sell. While civil liberty organisations still complain about the bias in algorithms, the public are wary – neither trusting, nor distrusting, but conscious that crime rose several years in a row with cutbacks on police officers. Predictive policing was touted as the artificial intelligence that was going to make huge cuts in crime – which, of course, has not happened as organised crime gangs have upped their game too.

Politicians, recognising the need to boost their trust with the public, agreed to adopt a new regulation making algorithms explainable to the public. Each algorithm was to include a bit of code saying who created the algorithm, who paid for it, its purpose, website and contact for more information. This dispelled concerns about the police wanting to keep their black boxes black, as it were, but led criminals to a better understanding of police methods and tactics and to a spate of hacking attacks on police systems. Meanwhile, some "grey hat" hackers attempted to improve the algorithms to help eliminate bias.

A significant factor in gaining public acceptance was the establishment of trusted independent national bodies to oversee police use of algorithms in predictive technologies. Adequately funded (for a change!) and staffed with known and respected figures such as Baroness Lawrence in the UK, these independent bodies helped to build trust in the police system. These bodies looked at not only the algorithms themselves, but all aspects of police use of data. They considered what data was collected, the purpose of its collection, how the data were processed and storied, and its eventual usage (including secondary use).

The findings of these bodies were, in the early days, significant in developing crucial training programmes for the police about the new technologies and their limitations. So new police officers are concerned about automation bias, regulations in 2025 spelled out what the police were permitted to do with data. Politicians and senior police officials communicated these rules effectively to the public. They hosted regular stakeholder engagement meetings with the public to ascertain their concerns. Local police forces have also been hosting local meetings with residents and community leaders to explain their use of new predictive policing technologies, how these technologies were vital in offsetting the cuts in police staff numbers and, importantly, how accurate these algorithms were in predicting criminal acts.

Do you think these actions are plausible and probable?

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Steps towards a desired future and avoidance of an undesired future

Civil society organisations, late night talk-show hosts and some editorial writers articulated their fears that the new predictive policing technologies would yield many false positives, and that perfectly innocent citizens could be victimised by the new technologies; that they could be placed on a police register without knowing why. There were worries about positive feedback loops in particular locales targeted for attention, leading to a greater number of arrests in these areas, and in turn to algorithms predicting that these were the areas on which the police should be concentrating. Had there been a blind trust in the efficacy of the algorithms, then this may well have been the case, but fortunately this concern had been raised so many times that the police and their algorithm developers were on guard for such phenomena.

By addressing these concerns directly, by instituting transparency measures and empowering oversight bodies, the police increased public trust and strengthened social cohesion. Predictive policing technologies helped the police to focus on areas of crime that were previously invisible. Data analysts uncovered these areas by training their PP algorithms with masses of information from disparate sources. This allowed the police to put more effort into tackling white collar crime and online hate crime. This in turn has had a ripple impact on international crimes such as people trafficking and drug smuggling. In fighting such crimes, the police noticed positive effects in communities that were otherwise subject to the attention of such smugglers. Overall, predictive policing has led to a decline in crimes. Criminals and their would-be accomplices now recognise that if they commit a crime, the likelihood of getting caught is higher than ever, even though there are lingering worries about the inevitability of at least some false positives which could lead to the imprisonment of innocent people.[110]

The police also appreciated the new technologies as they found that the effective intelligence led to their approaching volatile situations with an enhanced awareness of how those situations were likely to play out. These days, it’s rarely the case that a police officer finds himself unexpectedly in the middle of a riot and fearing for his life.

Predictive policing technologies have especially emphasised the prevention of crimes – not only by minutes or hours, but also on the factors that lead to criminality. The initial emphasis on street crime led to an outcry by CSOs, the media and citizens that such technologies were ignoring corporate crime which has a much bigger impact on society as a whole. Always loving a challenge, data scientists recently developed new smart information systems that are expected to significantly enhance the detection of corporate crime and questionable practices. These new technologies are

[110] Macnish, K., “Unblinking eyes: the ethics of automating surveillance”, *Ethics and Information Technology*, Vol 14, 2012, pp. 151–167. https://doi.org/10.1007/s10676-012-9291-0

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

bringing ethicists and data scientists together, which is expected to greatly benefit European competitiveness.

7. Recommendations for a desired future and avoiding an undesired future

From the above steps, we extract the following key recommendations to reach a desired future and avoid an undesired future:

- Clear and transparent criteria for personal data should be entered into law enforcement databases.
- Member States should have or establish an independent authority of sufficient size and clout to monitor the data in and use of law enforcement databases and offer commendations or impose penalties where appropriate.
- Measures in *preventive* policing and community investment should supplement developments in *predictive* policing.
- Law enforcement authorities should have a balanced approach to local, white-collar and online hate crimes and should not unduly emphasise street crime prevention at the expense of curtailing white-collar crime, for example.
- LEAs should offer more (effective) training of police officers and database operators as to the limitations of data analysis, particularly concerning rates of false positives.
- The EU should sponsor research on automatically detecting when an attack is being planned and discussed on criminal forums, and on predicting future threats.

Do you agree with these recommendations?
Are there any others that you think we should include?

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

6. Fourth scenario: Self-driving vehicles: navigating towards an ethical future

Image Credits: smoothgroover22, Flickr (CC BY-SA 2.0)

Figure 13 Fourth scenario - Image credits: smoothgroover22, Flickr (CC BY-SA 2.4)

1. Introduction

Self-driving vehicles (SDVs) offer great benefits for society, but also need to be carefully assessed and regulated before being integrated and used on our roads. In the following pages, we present a possible scenario for SDVs in 2025 in five key sections, concluding with what we can do in 2019 to ensure that we see a desirable future unfold, while avoiding some of the pitfalls outlined in the scenario. Our vignette discusses a München native, Adrian, and his use of an SDV in his hometown, portraying his use of the vehicle in 2025. This scenario also illustrates the main drivers and inhibitors that may affect the successful integration and adoption of SDVs between 2019 – 2025. The next section looks at possible ethical, legal, social and economic impacts of SDV use in the year 2025, which provides guidance on ways that these harms could be mitigated, while developing approaches that would accentuate the positive impacts of SDV use. The final section concludes with practical steps that we need to put in place now in order reach a desirable future for SDV use in the future.

In the next section, a vignette portrays a narrative of SDV use in 2025. The vignette serves the purpose to visualise how people will use this technology and how it will be adapted by 2025. This is followed by two sections to highlight potential driving forces behind the usage of SDV and what may be the inhibitors to its adaptation by 2025. SDVs are set to have a huge impact on our lives, so section 4 identifies social, ethical, legal and economic impacts; and how these may materialise by 2025. Following from this section, we review ways that we can mitigate the negative and accentuate the positive impacts of SDVs through policy implementation from 2019-2025. The concluding section will bring us back to the year 2019 in order to establish the steps we need to take towards a desired future and avoidance of an undesired future by 2025. The 2025 future outlined has desirable and undesirable features, which this section will try to address to provide recommendations. All of the scenario sections are written from the perspective of someone in 2025, except the final section on recommendations, which is written from the present.

Before leaping into the future, let us briefly consider the technologies that make SDVs possible.

SDV technologies

SDVs gained public recognition through the three DARPA challenges in 2004, 2005, and 2007[111]; which resulted in the establishment of four SDV characteristics: sensing, perception, planning and control. Sensors are used to take raw data measurements, which are transformed by the perception component into usable information. The planning component creates a path based on that information, and the control component contains the actuators to drive the car (based on the planned path, through direct sensing, in order to avoid obstacles)[112]. A combination of camera, radar and laser systems are used to retrieve data about the environment[113][114]. For the position and motion of the car, SDVs are equipped with satellite navigation, inertial and odometry measurements[115].

LiDAR (Light Detection and Ranging or Laser Imaging Detection and Ranging) lasers are the primary sensors used in environmental perception[116]. LiDAR technology works by rotating a laser sensor, providing several million data points per second, creating detailed maps of the nearby surroundings for detecting static and moving objects[117].

[111] Zhang, Xinya, Hongbo Gao, Mu Guo, Guopeng Li, Yuchao Liu, and Deyli Li. "A Study on Key Technologies of Unmanned Driving", *CAAI Transactions on Intelligence Technology,* Vol. 1, Issue 1, 2016, pp. 4-13.

[112] Campbell, Mark, Magnus Egerstedt, Jonathan P. How, Richard M. Murray, "Autonomous Driving in Urban Environments: Approaches, Lessons and Challenges", *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, Vol. 368, Issue 1928, 2010, pp. 4649-4672.

[113] Luettel, Thorsten, Michael Himmelsbach, and Hans-Joachim Wuensche, "Autonomous Ground Vehicles-Concepts and a Path to the Future", *Proceedings of the IEEE*, Vol. 100, Centennial-Issue, 2012, pp. 1831-1839.

[114] Zhao, Jianfeng, Bodong Liang, Qiuxia Chen, "The Key Technology Toward the Self-Driving Car", *International Journal of Intelligent Unmanned Systems*, Vol. 6, Issue 1, 2018, pp. 2-20.

[115] Campbell, 2010, p. 4654

[116] Hirz, Mario, and Bernhard Walzel, "Sensor and Object Recognition Technologies for Self-Driving Cars", *Computer-aided Design and Applications*, Vol. 15, Issue 4, 2018, pp. 501-508.

[117] Luettel 2012, p. 1833

Velodyne LiDAR is the best way to detect and track static and moving objects in urban traffic. By removing all ground points and performing 3D clustering on the remaining points, hypotheses can be made on the motion of objects[118]. Clusters can be classified into categories such as cars, bikes and pedestrians[119].

Great advancements are being made towards road perception by transforming road shapes and markings from 3D estimates into 2D images[120]. There have been developments in computer vision to interpret traffic lights and signals, but more work is required. Advancements in computer vision, combined with LiDAR technology, has enabled self-driving vehicles to overcome many issues related to poor performance at night, ambient lighting conditions, and bad weather conditions[121]. Developments in electronic mapping have aided the car's navigation by incorporating geographical characteristics, traffic information, building information, and traffic signs. The satellite navigation systems – GPS, GLONASS, Galileo, and Beidou – have spurred developing of electronic maps[122].

The navigation process of SDVs can be classified into four sections: route planning, behavioural decision-making, motion planner, and vehicle control[123]. The route planning stage involves selecting a specific route to the destination from digital map data. This is done by representing the road as a directed graph with edge weights corresponding to the cost of riding over a road segments, a suitable route is then found in the road network graph[124]. The behavioural component monitors traffic information and observes the behaviour of other vehicles in order to reach its destination[125]. Adapting to real-world uncertainties, and the intent of other traffic participants, has been one of the biggest challenges for SDVs; with developments in machine learning techniques, such as Gaussian mixture and regression models, enabling better traffic trajectory predictions[126].

There have been vast improvements in higher machine learning capabilities, with much of the cognitive automation being done with advanced deep learning and neural network-based models, such as recurrent, generative adversarial, and long-short term memory. There have been great developments used in hardware of the past number of years, allowing for faster processing of neural networks. Motion planning determines the best path for a car to take, comfortable for the

[118] Luettel 2012
[119] Luettel 2012
[120] Luettel 2012, p. 1834
[121] Hirz 2018, p. 6
[122] Boer 2017, p. 21
[123] Paden, Brian, Michael Čáp, Sze Zheng Yong, Dmitry Yershov, and Emilio Frazzoli, "A Survey of Motion Planning and Control Techniques for Self-Driving Urban Vehicles", *IEEE Transactions on intelligent vehicles*, Vol. 1, Issue 1, 2016, pp. 33-55.
[124] Paden 2016
[125] Paden, 2016, p. 5
[126] Paden 2016

passenger, while avoiding collision[127]. The trajectory calculated by the motion planning is performed by selecting the appropriate actuator inputs based on the planned movement, and the vehicle control tracks without a feedback controller loop system[128]. Early car-to-car developments use small radio transmitters and receivers on each car to broadcast information about location, speed and direction to other vehicles, to determine safe lane changed and merges[129].The DARPA Urban Challenges emphasised the importance for SDVs to access each other's information to effectively map trajectories, share their driving data, and update their digital maps[130]. The early developments of SDVs concentrated on self-contained vehicles, but it was not until they incorporated vehicle-to-vehicle communication (V2V), as well as vehicle-to-infrastructure communication (V2I), did they truly progress[131].

In the following section, we now jump ahead to the year 2025.

2. Vignette

In 2025, self-driving vehicles (SDVs) are used in different urban areas throughout the world. 38-year-old Software Developer Adrian uses his self-driving car to go to his office in München every morning, which was one of the first places to roll out SDVs. "So far, so good", explains Hans, who has been using his SDV for over 4 months now. "I am able to work in my car while commuting. When you factor in an hour commute each way, I get back 10 hours of my life that is lost in the commute every week. I sit back with my laptop, while listening to Spotify. It's great!". Hans' Waymo Centauri b is one of the few permitted self-driving car models on the market and has been one of the most widely adopted of these vehicles, so far.[132] The Centauri b is still in the hybridisation stage towards full automation, having both automated, semi-automated, and manual driving possibilities at level 4 automation.[133] Legally, Hans can

Do you think the use of a vignette helps to makes it easier for stakeholders to relate to the technology and its impacts?

[127] Paden 2016, p. 5

[128] Paden 2016

[129] Gora 2016, p. 2208

[130] Boer 2017, p. 21

[131] Lari, Douma, and Onyiah, 2015, p. 743

[132] The Waymo Centauri b was named after the closest habitable exoplanet in the solar system to represent the vehicle's similarity to our own, but which offers all kinds of wonderful hopes for humanity's future.

[133] There have been six distinct stages towards full automobile automation, starting at level 0 to level 5. Level 0 refers to automobiles that have no automation whatsoever, whereby the driver performs all actions and driving tasks. Level 1 refers to the driver assistance stage, whereby the vehicle is still controlled by the drive, but there are some features to assist the individual in their driving. Level 2 refers to partial automation, where there is driving automation in certain aspects of the driving experience, i.e., acceleration and steering. However, the driver needs to remain fully engaged throughout and take over if necessary. Level 3 refers to 'conditional automation', where more control is given to the automated vehicle, particularly in the monitoring of the environment, but the driver must still be ready to take over if there are any issues. Level 4 depicts high automation of the vehicle, where the vehicle has the capacity to respond to most aspects of the driving experience, leaving almost full disengagement of the driver. Level 5 is when the vehicle is 'capable of performing all driving functions under all conditions'. Please see: National Highway Traffic Safety Admin (NHTSA), "Automated Driving Systems: A Vision for Safety", *U.S. Department of Transportation* [website], September 2017, available here: https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13069a-ads2.0_090617_v9a_tag.pdf

only drive fully automated within designated areas of München, but for most other places the car must be in semi-automated or manual mode. “It is a nuisance when I have to drive outside München. It takes a while to get used to the wheel again. But I understand that it will take other cities time before they catch up with us,” Hans claims, as the vehicle navigates through his neighbourhood in level 4 automation. His car changes lanes and stops at pedestrian lights, gives way at roundabouts, while allowing him the comfort to catch up on work or just relax and take in the scenery. So far, SDVs have gained universal integration in only seven cities in the world, but there are hopes that this number will increase dramatically by 2030. Many of the leading car manufacturers and experts estimate that this number will be between 50 – 70 cities by the end of the decade. Some of the most pioneering and revolutionary developments have been coming from Silicon Valley, while the most prolific countries behind SDV development have been the US, South Korea, the UK, Japan, China, and Germany. The US has been the real innovator behind SDVs, with more than 40 cities piloting SDVs as far back as 2017, dwarfing all other countries in comparison[134]. At the start of 2025, there were 100 cities in the US piloting SDVs, and this number is set to increase dramatically by 2030. Hans has reaped the benefits of autonomous driving, but only after he passed his SDV driving test. In addition, cities integrating SDVs must also be authorised with the National Self-Driving Vehicle Transportation Board (NSDVTB) and the vehicle owner must be registered with the Department of Self-Driving Vehicles Authority (DSDVA). The vehicle itself must pass strict manufacturing standards before being allowed on the market. Outside of these designated areas, cars must function at level 3 capacity – limited automation. The car senses when conditions require the driver to retake control and provides a sufficient transition time for the driver to do so. Some SDV companies wanted to skip this stage, but the limitations of technological organisation, the interaction with manual drivers, and the lack of infrastructure to accommodate this move have been too problematic. In areas where there are mixed drivers (automation and non-automation), SDVs must have a level 3 option for legal reasons. One of the main reasons behind these laws is to ensure safety, which has been one of the main driving forces behind the development of SDVs in the first place.	

3. Drivers

Safety drivers: Approximately 90 percent of crashes are the result of mistakes by the driver and while road deaths have been decreasing,	Do you agree with these drivers? Are there any other significant drivers that

[134] Hao, Karen, “At least 47 cities around the world are piloting self-driving cars”, *Quartz* [website], December 4th, 2017, available at: https://qz.com/1146038/at-least-47-cities-around-the-world-are-piloting-self-driving-cars/

they were as high as 1.4 million in 2015[135] [136]. Over the past ten years, safety has been one of the strongest drivers for the implementation of SDVs, but we have yet to reap their true benefits because of their relatively low implementation. There is an ambitious goal to have zero automobile-related deaths in the United States by 2050, which may be feasible if they are successfully adopted nationally[137]. As far back as 2017, there have been studies to show that deploying SDVs when they are only marginally safer than humans (say, 10%), it would still have a dramatic impact on reducing road deaths. Policymakers around the world have largely indicated that waiting for SDVs to be far safer than humans (say, 75 – 95%) is not an option because of how long it would take to reach that stage[138]. *Social drivers:* One of the main drivers for SDVs has been that they would allow a greater diversity of people to drive, such as the blind, and some of the elderly and disabled population[139]. They may also offer people the ability to work, sleep, read, eat, or watch TV, while driving[140]. Because of the limited use of SDVs, we are yet to see a huge change in road efficiency and reduced traffic jams, which will require a sophisticated and intelligent transportation management systems to accommodate them. In 2014, the American Trucking Association (ATA) predicted that there would be a huge shortage of truck drivers, which would necessitate the development of self-driving trucks. Their prediction of 175,000 drivers by 2024 actually came up short of the reported 215,000-figure taken in November 2024[141]. SDV trucks have also shown promise to reduce carbon emissions through more fuel-efficient driving. *Environmental drivers:* In the cities where SDVs have been integrated, there has been an increase in public transport and car-sharing because of the novelty of being in a SDV, thus reducing overall carbon emissions, in addition to many SDVs being electrically powered. There has been a	should be included here?

[135] National Highway Traffic Safety Admin (NHTSA), "US Department of Transportation, Preliminary Statement of Policy Concerning Automated Vehicles", NHTSA Preliminary Statement, 2013, available at http://www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf

[136] World Health Organization, "The Top 10 Causes of Death", WHO [website], 24th May 2018, http://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death

[137] Ecola, Liisa, Steven W. Popper, Richard Silberglitt, and Laura Fraade-Blanar, "The Road to Zero: Executive Summary A Vision for Achieving Zero Roadway Deaths by 2050", *RAND Corporation* [website], 2018, available here: https://www.rand.org/pubs/research_reports/RR2333z1.html

[138] Kalra, Nidhi, and David G. Groves, "The Enemy of Good Estimating the Cost of Waiting for Nearly Perfect Automated Vehicles", *RAND Corporation* [website], 2017, available here: https://www.rand.org/pubs/research_reports/RR2150.html

[139] Lari, Adeel, Frank Douma, and Ify Onyiah, "Self-Driving Vehicles and Policy Implications: Current Status of Autonomous Vehicle Development and Minnesota Policy Implications", *Minn. J.L. Sci & Tech*, Vol. 16, Issue 2, 2015, pp. 735-769.

[140] Johnsen, Annika, Clemens Kraetsch, Katarina Možina, and Alba Rey, "D2. 1 Literature Review on the Acceptance and Road Safety, Ethical, Legal, Social and Economic Implications of Automated Vehicles", *BRAVE: BRidging Gaps for the Adoption of Automated Vehicles*, EC-Funded Project, No 723021, November 30th, 2017.

[141] Seattle Truck Law PLLC, "The Truck Driver Shortage Is Getting Worse — And More Dangerous", *Seattle Truck Law PLLC* [website], April 13th, 2018, available here: https://www.seattletrucklaw.com/blog/the-truck-driver-shortage-is-getting-worse-and-more-dangerous/

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

huge demand for more environmentally-sustainable vehicles since the Kyoto and Paris climate agreements. Cities view electric SDVs as one way to meet their EU carbon emission requirements[142]. However, there is still a concern that there will be an intersection between more commuting as a result of SDV convenience, leading to overall increased car usage. Since 2023, there have been a number of auto manufacturers testing single-user SDVs to bring people from their homes to public SDV buses, which would further reduce our environmental impact, while reducing costs for citizens. *Economic drivers:* While the price of SDVs is reducing every year, they are still more expensive than non-automated cars. Some people have proposed that SDVs could be shared in order to reduce costs, so that they do not sit idle in people's garages or parking lots and are be used throughout the day[143]. In addition, fuel costs for SDVs have been lower because of greater fuel-efficiency and when they reach widespread level 4 integration and safety is improved, it will reduce the necessity for airbags and steering wheels[144]. The whole design will change because of narrower, smaller and more economically viable vehicles[145]. Between the 2020 – 2025 period, a large number of new non-traditional players, such as ICT and data analytics companies, have emerged in the SDV automotive market. Many of the smaller automotive companies view SDVs as a threat because they cannot put the same kind of investments into developing these technologies as their larger automotive counterparts, which will lead to many foreclosing in the coming years as a result of market pressures. *Market drivers:* Over the years, some have stated that the SDV market is supply-driven and many people do not want to use them[146]. However, SDVs have witnessed development as a result of the need to transport goods, and businesses also view SDVs as another opportunity through the data retrieved from the vehicles[147][148]. Auto manufacturers have been hugely competitive in the race to develop SDVs, bringing global success and prestige to their company. Companies have been extensively patenting their cars, products, and services to lock	

[142] European Environment Agency, "Electric Vehicles in Europe", *EEA Report No 20/2016*, September 26th 2016, available here: https://www.eea.europa.eu/publications/electric-vehicles-in-europe

[143] Ohnsman, Alan, "The End Of Parking Lots As We Know Them: Designing For A Driverless Future", *Forbes*, May 18th 2018, available here: https://www.forbes.com/sites/alanohnsman/2018/05/18/end-of-parking-lot-autonomous-cars/#3aa6feac7244

[144] Davies, Alex, "GM Will Launch Robocars without Steering Wheels Next Year", *Wired*, January 12th 2018, available here: https://www.wired.com/story/gm-cruise-self-driving-car-launch-2019/

[145] Lari, Douma, and Onyiah 2015, p. 754

[146] McCarthy, Niall, "Global Opinion Divided on Self-Driving Cars", *Forbes*, April 13th 2018, available here: https://www.forbes.com/sites/niallmccarthy/2018/04/13/global-opinion-divided-on-self-driving-cars-infographic/#7d6eed5c110f

[147] DHL, "Self-driving Vehicles in Logistics", *Delivering Tomorrow*, 2014, available here: https://delivering-tomorrow.com/wp-content/uploads/2015/08/dhl_self_driving_vehicles.pdf

[148] Hawthorne-Castro, Jessica, "Autonomous Vehicles Will be a New Opportunity for Marketers", *Forbes*, June 4th 2018, available here: https://www.forbes.com/sites/forbesagencycouncil/2018/06/04/autonomous-vehicles-will-be-a-new-opportunity-for-marketers/#6b243a381b0b

customers into their brand. However, the notion of automotive branding has been changing over the past few years, with a shift from luxury, status and appearance, towards efficiency, safety, and functionality. *Efficiency and productivity drivers:* As a result of greater driving efficiency, SDVs opened up the possibility of reducing traffic jams and congestion, identifying better routes to take, driving more sustainably, and a reduction of crashes holding up traffic flow. Despite SDVs being heralded as a way where people can get extra sleeping or relaxation time on their commutes to work, some businesses view them as holding the possibility of cutting out needless 'driving time', so their staff can work while in the vehicle. *Political drivers:* Greater SDV driving efficiency is witnessing a reduction in lane size and quantity of lanes in areas restricted to level 4 automation[149]. Car-sharing has been increasing in these cities, which will eventually mean less public investment in parking lots because cars will be used throughout the day[150]. While the roll-out of SDVs is still new, cities will eventually witness a reduction in healthcare spending because of fewer automobile accidents. SDVs have been reducing the need to live in urban areas for work because people are able to commute from farther away without the strains of previous commutes, relieving resource strain on these areas[151].	
4. Barriers and inhibitors *Safety and security barriers:* Many different safety issues slowed down development of SDVs. For example, the motion sickness of passengers is an issue SDV developers have been trying to solve for the past decade[152]. Overall, the safety of automated vehicles has been a primary concern amongst road-users, especially following some of the highly-publicised deaths, such as the Tesla Model S in 2016[153]. People have found it difficult to put their safety in the hands of an autonomous machine for fear of technical or systems failures, malfunctions, or just general unreliability of these new vehicles[154]. While crashes with SDVs have decreased over the past few years, they are still more risk-prone in terms of accidents per mile driven than driver-controlled vehicles. Even going back as far as 2017, figures indicated accident rates for every 48,000 miles driven for SDVs, compared to every 2.08 million miles driven for non-autonomous	Do you agree that these are likely to be the most significant barriers and inhibitors in 2025? Are there any other barriers that should be mentioned?

[149] Lari, Douma, and Onyiah 2015, p. 752

[150] Ibid., p. 758

[151] Ibid., p. 757

[152] Sivak, Michael, and Brandon Schoettle, "Motion Sickness in Self-driving Vehicles", *The University of Michigan Transportation Research Institute*, April 2015, available here: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1001.6169&rep=rep1&type=pdf

[153] Stilgoe, Jack, "Machine Learning, Social Learning and the Governance of Self-driving Cars", *Social Studies of Science*, Vol. 48, Issue 1, 2018, available here: http://journals.sagepub.com/doi/abs/10.1177/0306312717741687

[154] Johnsen et al. 2017, p. 21

cars[155]. Mode transitions has raised additional safety issues, such as distraction, loss of situational awareness, and high workload during take-over. All of these factors have proven to be inhibitors to the successful development of SDVs and are issues that are constantly being tested and rectified. Many people have also been worried about the security risk of SDVs, such as hacking, manipulation and malicious activity.

Technical barriers: There have been many technological barriers to SDV development, including issues around data security, vehicle security, hacking and cyber-security. Initially, steering systems had built-in processes to determine abnormal instructions, but after a few minor concerns relating to compromised commands, SDVs were implemented with a 'master computer' that takes control and brings to vehicle to a safe stopping position in the case of suspicious activity. Auto manufacturers have been pressured to provide increased AI transparency, which has inhibited the speed of development, as have the challenges of ensuring sufficient software and hardware updates. Locations need to have 5G technology access, which has been a limiting factor to SDV integration in many places[156]. Vehicles request relevant information about their current position from the cloud, overcoming the limitations of sensor-based information[157]. Both automotive and ICT companies have also had to invest heavily into their frequency communication infrastructure as there was an unwillingness by governments to finance these systems at the speed required to facilitate SDV integration.

Political barriers: Since SDVs were first developed, there has been a difficulty to establish standardisation between companies and countries. It has been challenging to develop protocols, with some claiming that regulation has been too stringent, halting progress, while others have stated that it has not been stringent enough. Governments have found it difficult to strike an appropriate balance between the two and there has also been a great deal of diversity with SDV policies globally, ranging from extremely detailed and dense (EU, US, and Japan) to non-existent (Eritrea, North Korea, and Somalia).

Economic and geographic barriers: One barrier for SDVs adoption has been their cost and the infrastructure required to facilitate them. It has been costly to implement policies, procedures, and technical arrangements to accommodate SDVs, so they have largely been adopted by wealthier countries. They have mostly remained untested in many of the world's poorer countries, which is proving to be a key concern in global SDV and social justice circles. Even within richer nations, there has been a wide divergence in acceptance rates of SDVs.

[155] Ibid., p. 33

[156] Boer, Arnoud, Rob van de Velde, and Marc de Vries, *Self-Driving Vehicles (SDVs) & Geo-Information,* 2017, retrieved from https://www.geonovum.nl/uploads/documents/Self-DrivingVehiclesReport.pdf

[157] Kumar, Swarun, Shyamnath Gollakota, and Dina Katabi, "A Cloud-Assisted Design for Autonomous Driving", in *Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing*, 2012, pp. 41-46.

For example, willingness-to-pay studies have varied widely amongst nationalities, with many of these divergences remained unchanged since 2017, despite national efforts: ‘Italian participants were most interested in using autonomous vehicles (65 %), followed by the Spanish participants (54 %), the French participants (51 %), the Belgian participants (50 %), the German participants (44 %) and the American participants (32 %)’[158]. Location has played a fundamental role in the acceptance or rejection of SDVs, due to varying local attitudes, reliance on employment in driving professions, and technological capabilities, as well as economic stability of the country. For example, despite there being a greater acceptance rate among Italian and Spanish citizens, the economic instability of both regions over the past decade has inhibited the integration of SDVs.

Employment barriers: One of the main inhibitors to the acceptance of SDVs has been a concern around job security. There has been an increased concern in recent years about SDVs replacing taxi drivers, bus drivers, delivery drivers, and anyone dependent on driving as a profession. Many trades unions and organised workforces in these areas have petitioned and protested at the replacement of workers in these sectors. Animosity towards SDVs from these groups has led to isolated incidences of abuse towards SDV taxi managers, destruction of vehicles and protests outside Waymo headquarters in Mountain View.

Social barriers: There has been a lot of negative publicity about SDVs, especially about fatalities, such as Uber’s accident in 2018. There have been many cases of local residents harassing SDV drivers, slashing tyres on vehicles, throwing rocks, and hostility towards them[159]. The media has sometimes been criticised for focusing on many of the negative aspects of SDVs, such as the crashes and fatalities, which has affected public understanding and acceptance of the vehicles. Providing a level of trust amongst the public in relation to crashes, hacks and malfunctions has been one of the greatest challenges for SDVs market integration. Because SDVs are relatively new to the market, it has also been difficult to estimate user acceptance. In many reports, there is an expressed fear that others will have access to your data. Some organisations have even established protocols to ensure that users’ privacy is protected when selling their SDVs[160].

Data protection and privacy barriers: Since the creation of the GDPR and the many controversial data leaks and privacy debacles over the past seven years, there has been a heightened concern about data protection and privacy, which has inhibited SDV deployment. SDV developers have been trying to navigate between privacy and data protection on the one side, and the need for vast amounts of

[158] Johnsen et al. 2017, p. 25

[159] Cuthbertson, Anthony, “People are Slashing Tyres and Throwing Rocks at Self-driving Cars in Arizona”, *Independent Newspaper* [website], December 13th, 2018, available here: https://www.independent.co.uk/life-style/gadgets-and-tech/news/self-driving-cars-waymo-arizona-chandler-vandalism-tyre-slashing-rocks-a8681806.html

[160] NADA [National Automobile Dealers Association], “Personal Data in your Car”, *NADA* [website], 2018, available here: https://www.nada.org/PersonalDataInYourCar/

processing data for SDVs to function, on the other. After the first large fine of €50 million against Google back in 2019 from European regulators created a snowball of large ICTs being heavily fined, there has been a strong fear in the industry about breaching the GDPR[161]. The GDPR has sometimes proven to be a hurdle for SDV manufacturers selling into the EU market; whereas, countries not abiding by this regulation have been able to develop a little quicker. Overall, there has been a global concern about what SDV manufacturers can do with SDV data without infringing on individuals' privacy and abiding by the legalities of the GDPR.

Legal barriers: There has been a difficulty uniting cohesive legal analysis due to national differences on road traffic and transportation. One of these barriers has been determining accountability in cases of accidents. Manufacturers have tried to keep accountability in the hands of the driver, keeping SDVs at level 3 automation. However, this has also prompted some manufacturers to take full responsibility in order to promote trust in their vehicles. The different levels of accountability have led to some confusion in the insurance industry about how to deal with accidents.

5. Ethical, legal, social and economic impacts

Ethical Impacts

Safety and prevention of harm: In discussions of SDVs, one sometimes hears questions about whether non-automated driving should be banned when we reach a level where SDVs can safely and easily replace non-autonomous driving. While still in the hypothetical stage, once SDVs become prevalent, 'it seems morally or ethically necessary to prohibit selling and using non-autonomous vehicles'[162]. Because the rollout of SDVs has been so slow, this has not been a pressing question, thus far. Meanwhile, groups such as Humans Against Autonomous Vehicles (HAAV) have strongly opposed SDVs because they are not safe enough to drive and are just "glorified smartphones".

Do you agree that the ethical issues listed here are likely to be important in 2025?

Are there any other ethical issues we should include?

Another concern is what an SDV should do if there is an unavoidable crash: How should the SDV be programmed and who should determine these priorities[163]. Nobody would buy an SDV if they prioritised the lives of others over the vehicle's driver and passengers[164]. However, if algorithms aim to protect the driver, they may crash into children or light vehicles, instead of other cars, walls, or lampposts, to protect the

[161] Porter, Jon, "Google Fined €50 Million for GDPR Violation in France, *The Verge* [website], January 21st 2019, available here: https://www.theverge.com/2019/1/21/18191591/google-gdpr-fine-50-million-euros-data-consent-cnil

[162] Johnsen et al. 2017, p. 39 and also Nyholm, Sven, "The ethics of crashing with self-driving cars: A roadmap, II", *Philosophy Compass,* Vol 13, Issue 17, 2018.

[163] Contissa, Giuseppe, Francesca Lagioia, and Giovanni Sartor, "The Ethical Knob: ethically-customisable automated vehicles and the law", *Artificial Intelligence and Law,* Vol. 25, Issue 3, 2017, pp. 365-378.

[164] Johnsen et al. 2017, p. 3

driver[165]. Also, if safety is the main concern, they may swerve towards a motorcyclist wearing a helmet, as opposed to one without a helmet, because they would be more likely to survive in a crash[166][167]. If algorithms target those less at risk, then people may start to take unsafe activities in order to become safe, i.e. cycling without a helmet so that SDVs view you cautiously, thus avoiding collision[168]. Manufacturers want to confine SDVs at level 4 to areas that prohibit non-autonomous vehicles, because the uncertainty of non-autonomous driving is the biggest risk. *Moral algorithms:* Algorithms determine statistical likelihoods that certain groups of people would be more likely to die in a collision[169][170]. Surveys to identify driving behaviour are inaccurate because some people feel pressured to give more self-sacrificing, altruistic answers, than they would in reality. However, it is naïve to assume that people are generally self-sacrificing in split-second decisions, which has been verified in driving simulations and experiments[171]. Therefore, creating crash algorithms based on social values, or even individual values, is difficult to incorporate within SDV driving algorithms. While there have been guidelines and recommendations, regulation is still not fundamentally clear for SDV programmers. *Autonomy:* There has been a concern that in specific life-or-death scenarios, programmed responses may remove control from the human being in specific circumstances. We lose the choice and ability to make split-second decisions that could imperil our lives or those around us, but that we should make these decisions, otherwise it hinders our autonomy[172]. Car manufacturers have been concerned about how to program SDVs in specific scenarios, because these pre-given responses may not correspond to how we would behave in reality. Advocacy groups have claimed that car manufacturers will program the 'correct' response or that they may be forced to do so by regulation, which some propose diminishes human autonomy[173]. There has also been a concern that SDVs are threatening our free will and responsibility, because of the removal of accountability from the	

[165] Contissa, Lagioia and Sartor 2017, p. 67

[166] De Sio, Filippo Santoni, "Killing by autonomous vehicles and the legal doctrine of necessity", *Ethical Theory and Moral Practice,* Vol. 20, Issue 2, 2017, pp. 411-429.

[167] Johnsen et al. 2017, p. 42

[168] Ibid., p. 42

[169] Ibid., p. 43

[170] Nyholm, Sven, and Jilles Smids, "The ethics of accident-algorithms for self-driving cars: an applied trolley problem?", *Ethical theory and moral practice,* Vol. 19, Issue 5, 2016, pp. 1275-1289.

[171] Sato, Toshihisa, Motoyuki Akamatsu, Toru Shibata, Shingo Matsumoto, Naoki Hatakeyama, Kazunori Hayama, "Predicting Driver Behavior Using Field Experiment Data and Driving Simulator Experiment Data: Assessing Impact of Elimination of Stop Regulation at Railway Crossings", *International Journal of Vehicular Technology*, Volume 2013, available here: https://www.hindawi.com/journals/ijvt/2013/912860/

[172] Federal Ministry of Transport and Digital Infrastructure, "Ethics Commission: Automated and Connected Driving", *BMVI* [website], June 2017.

[173] FMTDI 2017, p. 16

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

individual as a result of overreliance on algorithms and artificial intelligence[174]. Already, in cities where level 4 automation is in place, there has been personal accounts of individuals feeling a loss of control in these vehicles. In other instances, there have been issues relating to lost control because SDVs have been programmed to abide by speed limits and rules of the road. For instance, in California recently, a pregnant woman went into labour and had to be rushed to hospital but was delayed because of the SDV's speed limit regulation.

Rights: There have been many rights-based issues related to SDVs over the past few years, some of which are still only hypothetical, such as: if a car is shared, who owns the car and what rights do you have to it? Policymakers have identified that while SDVs open the possibility for more people to use them than non-autonomous cars, it also poses the challenge of who do you deny the right to use them. As of now, countries are still following non-autonomous driving policies in relation to capacity to drive a car, as most still require level 3 automation. The elderly, blind and disabled are still being disadvantaged, but once SDVs reach widespread level 4 and 5 automation, they will begin benefitting from them.

Insurance and discrimination: There are concerns that SDV data will be used against individuals, and groups of individuals, by insurance companies. Now that cars are able to retrieve a wide array of driving habits, patterns, and behaviours, it means that if insurance companies gain access to this information, which many have already proposed an interest in, insurance could be tailored to meet individuals' driving performance. While being heralded as a positive move towards providing better insurance premiums to safer drivers, others have proposed that it would infringe on people's sense of privacy, with the feeling of constantly being monitored in the vehicle. Others have disavowed it because of the imbalance in insurance between manual cars and SDVs – namely, that insurance companies will provide better conditions for SDV drivers who allow their data to be monitored by insurance companies, to the disadvantage of non-SDV drivers.

Privacy: Privacy has been one of the most fundamental issues concerning the use and implementation of SDVs over the past decade. As a result of the large amounts of data retrieved from SDVs, policymakers need to identify methods to ensure privacy and data security; determine who should have access to this data; how it should be securely stored; and if law enforcement should be allowed to hack an SDV if it is breaking the law. So far, regulators have determined that strong levels of encryption, anonymization and aggregation need to be implemented in order to protect the individual's personal data. A lot of automobile manufacturers are promoting their DRIC compliant "data

[174] CNIL, "How can Humans Keep the Upper Hand? The Ethical Matters Raised by Algorithms and Artificial Intelligence", *Report on the Public Debate Led by the French Data Protection Authority (CNIL) as Part of the Ethical Discussion Assignment Set by the Digital Republic Bill*, December 2017, available here: https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf

remains in car" approach[175], which attempt to process and integrate data within the car, rather than being transmitted to different service providers or third-parties. This has been recommended since late 2024, but manufacturers have found it technically challenging to abide by.	
Legal impacts *Data and privacy:* SDVs produce huge amounts of data and require large processing capabilities. The massive amounts of data required to operate SDVs have raised privacy concerns about if individuals are identifiable, who has access to this data, and what can be done with it[176]. There has also been debate over whether data acquired from SDVs can be used as legal evidence; for example, if the driver was in control of the car at the time of an accident, could that evidence be used in court to determine liability[177]. Furthermore, concerns have been raised about how long data should be stored; where it should be stored (e.g., on the car's hard drive, the manufacturer's cloud platform or an independent cloud platform); who should be granted access to this data; under what conditions; what happens to the owners data when they sell the car; how will the data be protected from being hacked; and who *owns* this data[178]. Sensors collect information about the environment, which could be an infringement on bystanders' privacy. Because car companies are compiling mixed data (both personal and non-personal), it has been a little unclear how they are abiding by the GDPR. In addition, they have also had to incorporate how they were securely and safely protecting privacy in accordance with ePrivacy Regulations (ePR) created to ensure that automotive companies abide by its guidelines. The European Automobile Manufacturers' Association (ACEA) and the Council of the European Union have been paramount for ensuring that these governments implement the ePR and that those working in the industry follow the recommendations outlined[179]. *Cyber-security:* People have been fearful that SDVs will be easily hacked because of the abundance of digital infrastructure required for them to work. Criminals could make explicit use of the data that they retrieve, hack the vehicle and get it to perform actions the user is unaware of, unable to undo, or maliciously cause harm to the individual(s) in the car[180]. If cyber-criminals take over a vehicle, they may cause a nuisance	Do you think these will be the key legal issues in 2025? Are there any other legal issues that we should include?

[175] CNIL, "Connected Vehicles: A Compliance Package for a Responsible Use of Data", *CNIL* [website], February 13th, 2018, available here: https://www.cnil.fr/en/connected-vehicles-compliance-package-responsible-use-data

[176] Gogoll, Jan, and Julian F. Müller, "Autonomous cars: in favor of a mandatory ethics settings", *Science and engineering ethics,* Vol. 23, Issue 3, 2017, pp. 681-700.

[177] Johnsen et al. 2017, p. 53

[178] Johnsen et al. 2017, p. 53

[179] ACEA, "Proposal for ePrivacy Regulation", *ACEA* [website], June 4th, 2018, available here: https://www.acea.be/news/article/proposal-for-eprivacy-regulation

[180] Bowles, Jill, "Autonomous Vehicles and the Threat of Hacking", *CPO Magazine*, October 1st 2018, available here: https://www.cpomagazine.com/2018/10/01/autonomous-vehicles-and-the-threat-of-hacking/

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

with opening and closing windows or other minor grievances or even disable the car’s functionality to read stop signs, maliciously cause the vehicle to crash and harm its passengers or use the SDVs for terrorist purposes to transport remote-controlled bombs. While there is a greater need for transparency from car manufacturers, there is the problem that cars will become more vulnerable as a result. So far, there have been only a few minor issues related to cyber-security, such as the case in London where attackers found weaknesses in the SDVs through crypto malware and were able to extort money from the passengers before releasing control of the vehicle. However, these were isolated incidences and most of the cybersecurity insecurities have been identified by grey-hat hackers before malicious incidences occurred.

There has been a greater emphasis on strengthening counter-measures to avoid these situations. For example, in January 2025, UK police were granted the ability to take over cars that are hacked or under control for malicious purposes. This was done through the use of Decentralised Environmental Notification Messages (DENM), which are messages exchanged between peer-to-peer SDVs and their digital infrastructures[181]. DENM sends messages to the police if there are abnormalities, that indicate that the vehicle has been hacked, and comprises cryptographic signatures, which ensure that the messages being received from the SDV is from a reliable source, through certification and Public Key Infrastructure (PKI) architecture[182]. The certificates are linked with the vehicle at precise times and if the vehicle can be trusted. These anomaly-based detection methods are able to identify a lot of attacks, but miss others, so there have been developments towards remote attestation methods, which check protocols before granting access to services[183]. If there are abnormal issues addressed during this process, that indicate potential hacking, this is relayed to the Police ICT Departments for further testing before intervention.

*Liability**:*** Many motorists have been concerned about identifying liability in SDV crashes[184]. At levels 0-2, it is very clear that, legally, the driver is completely responsible for the car’s behaviour. SDVs become an issue at levels 3 and 4, because of the uncertainty of who is liable in cases of accidents. It is very important, under law, to identify who is responsible for the vehicle and under what circumstances[185]. So far, some traditional insurance companies have established insurance policies for SDVs, with premiums at the same rate as non-autonomous vehicles, unless the driver grants them access to their SDV data. SDVs raise the issue of who should be held accountable in case of accidents

[181] Article 29 Data Protection Working Party, “Opinion 03/2017 on Processing Personal Data in the Context of Cooperative Intelligent Transport Systems (C-ITS)”, *European Commission Data Protection* [website], adopted on October 4th, 2017, available here: http://ec.europa.eu/newsroom/just/document.cfm?doc_id=47888

[182] Article 29 Data Protection Working Party 2017, p. 4

[183] Kylänpää, Markuu, “Remote Attestation Adds Trust to Critical Infrastructures”, *VTT Blog* [website], November 10th, 2017, available here: https://vttblog.com/2017/11/10/remote-attestation-adds-trust-to-critical-infrastructures/

[184] Johnsen et al. 2017, p. 22

[185] Johnsen et al. 2017, p. 51

and thus responsible for compensation[186]. Since 2020, some of the main issues relating to SDV liability are: ● Determining accident liability if the driver is allowed to 'focus attention on tasks other than driving'[187][188]. For example, if the car is in self-driving mode and the driver is reading, but needs to quickly take control of the wheel, and fail to do so in time, should the driver be held accountable? So far, manufacturers have largely claimed responsibility for crashes at level 4, but at level 3, drivers are not permitted to do other activities that would prevent them from taking control of the wheel. ● Another problem relates to determining liability at the 'origin of the malfunction'[189]. It has been difficult to identify the point at which an SDV malfunctions to a specific time, making it a challenge to identify liability. Manufacturers have accepted responsibility for most cases of malfunction in recent years. ● If the driver activates the car when they should not have, or they do not take over control when requested, has also been an issue for liability detection[190]. For example, there have been cases where the driver is aware that they will be liable for an accident if they take control of the wheel, so they don't, thereby placing liability on the SDV's system. ● Problems arise when there is a critical situation and the driver and car react at the same time[191]. For example, a car driving in front of the SDV brakes and the driver turns the wheel to the right to avoid a collision, while the SDV veers the wheel to the left to avoid the collision. Both actions counteract one another and the car crashes into the back of the vehicle in front. For example, there was a situation in Seoul last year (May 6th, 2024), where this happened to an SDV driver. Luckily, nobody was badly injured, and the manufacturer admitted responsibility after reviewing the driver's inboard footage. ● Problems have occurred when there is a crash and the driver's response would have been better than the SDV's reaction time in the same situation[192]. For example, in one of the Volvo SDV test-runs in December 2021, there was thick fog on the road, impeding visibility. A cow came onto the road, but the SDV was unable to detect it and reacted too late and the car skidded out of control into a bollard nearby. The driver would have seen the	

[186] Lari, Douma, and Onyiah 2015, p. 760

[187] Gurney, Jeffrey, "Sue my car not me: Products liability and accidents involving autonomous vehicles", *Journal of Law, Technology and Policy,* Issue 2, 2013, pp. 247-277.

[188] Johnsen et al. 2017, p. 51

[189] Johnsen et al. 2017, p. 51

[190] Johnsen et al. 2017, p. 51

[191] Johnsen et al. 2017, p. 51

[192] Ibid., p. 51

cow earlier, with the use of his high-beam fog lights and could have avoided the crash.

- Concerns surround SDVs that break the law, when the driver is not required to monitor its actions[193]. It is difficult to determine if the driver is liable because they should have been monitoring the vehicle, or if the manufacturer is responsible because they implemented the SDV functionality that would break the law. So far, in the locations where level 4 vehicles have been integrated, manufacturers state that they are strictly following local laws and rules of the road, so this issue has yet to materialise in reality.

Social impacts

Joy of driving: For many, SDVs take away one of the primary pleasures of vehicles – the joy of driving itself[194]. While for some driving is a necessary ordeal that must be endured, for others, it is a form of pleasure in itself: a sense of control, a form of relaxation, a sense of adventure, and a connectedness with their surroundings, that is being threatened by SDVs. Some groups of driving enthusiasts are setting up affiliations to ensure that SDVs do not engulf their ability to drive in the future, but many say that the death of non-autonomous cars is an inevitability.

Figure 14 The joy of driving

Gender differences: Many years ago, the BRAVE Project was one of the first to highlight that there are different perceptions about SDVs

Do you think these are likely to be important social impacts created by these technologies in 2025?

Are there any others you think we ought to include?

[193] Ibid., p. 51

[194] Kemp, James, "Driverless Cars Will Take the Fun Out of Driving", *DriveWrite Automotive Magazine*, 2018, available here: http://www.drivewrite.co.uk/driverless-cars-will-take-fun-driving/

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

between men and women. Men have had less worry about embracing the technology, while women have been less enthusiastic and more fearful about the safety of SDVs and the difficulty of their use[195]. Male drivers showed a more favourable attitude towards SDVs[196]. Men have been buying SDVs at a greater rate than women, with an approximate 60-40 split in SDV usage. Manufacturers are supporting further research to determine how to increase female acceptability of SDVs.

Inclusion: SDVs hold the potential to reduce inequalities and promote inclusion amongst drivers by allowing certain groups (senior citizens, non-drivers, people with disabilities) access to automobiles that was limited, or unavailable previously[197]. However, because of the low levels of automation, this has not been possible, although many of these groups have indirectly benefitted from the use of SDV ride-hailing.

Car-sharing: While SDV car-sharing has not yet materialised because of low levels of automation, they hold the possibility of changing the nature of car ownership in the future. Some propose that SDVs will not remain static in garages or parking lots but will be shared amongst groups of people and used throughout the day, when we get to widespread level 4 and 5 automation[198]. Google's Waymo has been pioneering SDVs ride-hailing as far back as 2018 and have since introduced preliminary pilots in a number of cities throughout the US[199]. There were a few incidences in 2023, where passengers were not allowed to leave the car because of a glitch in the payment system, but overall, they have been a huge success and are set to expand their ride-hailing globally.

Travel behaviour and demands: It is still unclear if total travel miles increase as a result of 'travel comfort, convenience, and possibilities for non-drivers to use cars'[200]. So far, the limited integration of SDVs indicates that people travel more often as it eases many of the stresses found compared with traditional driving. In addition, fuel costs have been decreasing in five of the seven cities where level 4 automation has been implemented, because of more efficient driving, while the other two cities showed no change. In the past, it was assumed that insurance costs for SDVs would decline with a lower number of accidents. However, insurance companies are still dubious about the safety of SDVs and have kept insurance costs mostly the same as for non-autonomous vehicles, unless drivers can prove their safe driving through their SDV data. While SDVs initially had a higher number of accidents per mile than traditional cars, this was simply because they

[195] Johnsen et al. 2017, p. 28

[196] Ibid., p. 27

[197] Ibid., p. 56

[198] Ibid., p. 55

[199] Griswold, Alison, "Waymo is Readying a Ride-hailing Service that could Directly Compete with Uber", Quartz [website], February 16th, 2018, available here: https://qz.com/1208897/alphabets-waymo-googl-is-readying-a-ride-hailing-service-in-arizona-that-could-directly-compete-with-uber/

[200] Johnsen et al. 2017, p. 56

were in such early stages of development. Since July 2024, there has not been a fatal accident as a result of SDVs.

Efficiency: SDVs are allowing for closer travel proximity on the road from safer driving, while producing a more efficient traffic flow[201][202]. With the prospect of sharing SDVs, it may lead to less parking spaces if they are used throughout the day[203].

Decreased urbanisation: What has been happening in some of the cities where SDVs are being used is that drivers are beginning to live further away from the city centres because of the ease of commuting and reduced costs of running their SDV. There is less of a need to live in cities, which has started to see a reduction in urbanisation, allowing for a more evenly spread out population throughout the region. It has started to take some of the strains off amenities and busyness of very congested cities[204].

Environmental: There is an uncertainty about whether SDVs are ameliorating or exacerbating congestion levels. So far, people with SDVs have increased their overall travel time because they see it as less of a burden. However, it has been proposed that SDVs will improve efficiency and reduce congestion levels[205]. Early signs indicate that increased efficiency will reduce harmful carbon emissions more than non-autonomous vehicles[206]. SDV developers have been trying to walk the tightrope between ensuring their vehicles are environmentally-sustainable and having economically-affordable vehicles. Some manufacturers have placed a greater emphasis on emission reductions with the foresight that governments are implementing harsher penalties for poorly performing vehicles.

Economic impacts

Job-losses: In the past, there were worries that SDVs would lead to job losses for 'taxi drivers, parking attendants, valet parkers, car mechanics, meter attendants, traffic officers, and potentially bus and freight drivers'[207]. There have also been concerns that there were not enough people to drive trucks in places such as Canada[208]. Many truck manufacturers, such as Mercedes, noticed this trend and capitalised on autonomous trucks[209], and have been testing level 5 trucks for locations

Do you agree with these economic impacts?

Are there any other security and economic impacts that you think

[201] Gogoll and Müller 2016, p. 685

[202] Johnsen et al. 2017, p. 58

[203] Ibid., p. 59

[204] Lubell, Sam, "Here's How Self-driving Cars Will Transform Your City", *Wired*, 21st October 2016, available here: https://www.wired.com/2016/10/heres-self-driving-cars-will-transform-city/

[205] Johnsen et al. 2017, p. 58

[206] Gogoll and Müller 2016, p. 685

[207] Lari, Douma, and Onyiah 2015, p. 758

[208] CBC News, "Trucking Industry Facing Driver Shortage", *CBC* [website], July 15th, 2018, https://www.cbc.ca/news/canada/ottawa/trucking-shortage-ottawa-drivers-1.4746433

[209] Mercedes-Benz, "The Long-haul Truck of the Future", *Mercedes-Benz* [website], 2018, available here: https://www.mercedes-benz.com/en/mercedes-benz/innovation/the-long-haul-truck-of-the-future/

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

where it is too dangerous or unsuitable for humans to drive, since Autumn 2023. Uber saw that many of its drivers could become unemployed because of SDVs, so they have created computer science, engineering, and maintenance programmes for those interested in upskilling and transitioning professions[210]. *Competition:* As a result of the large investments and technological capacities of SDV development, we have seen a number of smaller automotive companies beginning to dissolve because they will be unable to compete with these giants going forward. While SDV start-ups flourished in the early infancy stage, the larger players have started outcompeting them with innovation, thus minimising the competitive market of SDV manufacturers. *Luxury vehicle business:* Some of the luxury vehicle manufacturers were worried about how SDVs would impact their business models, especially if driving were relegated to a hobby. However, some manufacturers have flourished through this period, with Audi and Mercedes taking leading roles in the SDV market[211]. However, companies such as Ferrari, Lamborghini and Lexus are trying to re-market their vehicles and have begun investing in 'drive for fun' initiatives and racing tracks. *Digital divide:* SDVs are very expensive, which has limited ownership to rich people[212]. It is difficult for poor people to drive SDVs and may become problematic when it becomes the prevalent form of transportation. There are concerns that the increased safety of SDVs may cause non-SDVs to be seen as unsafe and eventually prohibited from being sold, limiting people to more expensive SDVs. *Cost reduction:* In the past, it was suggested that SDVs would cause insurance and energy costs to decrease, but we have only witnessed minor changes.[213] While SDVs are hailed as safer, which should have reduced insurance costs, has not materialised in practice. *Tax and ownership:* One recent concern is related to the ownership of SDVs and who will be responsible for the taxation, insurance and maintenance of the vehicle, if they are shared. There have been developments in models of car ownership, with some companies, such as Uber, beginning to implement pay-as-you-use ownership models.	will be particularly important in 2025?

[210] Engelbert, Cathy, "Driverless Cars and Trucks Don't Mean Mass Unemployment – They Mean New Kinds of Jobs", *Quartz* [website], 1 Aug 2017. https://www.mercedes-benz.com/en/mercedes-benz/innovation/the-long-haul-truck-of-the-future/

[211] Autotech, "46 Corporations Working on Autonomous Vehicles", *CB Insights*, 4 Sept 2018. https://www.cbinsights.com/research/autonomous-driverless-vehicles-corporations-list/

[212] Oliver, Nick, Kristina Potočnik and Thomas Calvard, "To Make Self-driving Cars Safe, We Also Need Better Roads and Infrastructure", *Harvard Business Review*, 14 Aug 2018. https://hbr.org/2018/08/to-make-self-driving-cars-safe-we-also-need-better-roads-and-infrastructure

[213] Johnsen et al. 2017, p. 23.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Road infrastructure: There has been a lot of debate over whether governments should maintain existing infrastructure or start implementing a more digitised infrastructure to accommodate for SDVs[214]. So far, SDVs have had to develop to understand human signs, rather than digital signs. Furthermore, there has been a public outcry about governmental investment in SDV infrastructure, with many claiming that it should be partly funded by auto companies. In late 2024, demonstrations in France and Germany called on SDV manufacturers to aid cities pay for SDV infrastructure.

Law enforcement income: There has been a concern, in London and Mountain View, California, that SDVs will impact income generation of law enforcement. With more law-abiding vehicles, there has been a marginal and slow reduction in speeding and illegal parking. While more law-abiding vehicles is obviously a good thing, it still means a lost form of revenue generation by the police[215].

Electricity and power: While SDVs have been powered by a mix of electric and traditional fossil fuel, there has been a strong emphasis from governments to switch to electric. For example, the UK government stated back in 2018 that more than half of all vehicles on the road should be electric by 2030[216]. SDVs burn less fuel because of more efficient driving. However, in the cities where there have been large rollouts of SDVs, there has been an overall increase in fuel use because of their increased ease of use.

6. Mitigating the negative and acting on the positive impacts

As far back as 2019, there have been many different actions to mitigate negative impacts, while accentuating the positive impacts, of SDV technology, through national, international and supranational legislation and policy. One of the ways this was achieved was through national standardisation protocols between policy-makers, auto manufacturers, computer scientists, and transportation agencies. Standardisations have been created to ensure sufficient cyber security capabilities for SDVs are developed and implemented; minimum requirements established for the use of sensor technology; safety levels have been incorporated into earlier vehicle regulations to include hardware standardisations; and there have been several layers of enforced testing for different levels of vehicle automation.

Do you think these actions are plausible and probable?

[214] Peters, Mary, "Self-driving Cars Should Help Pay to Pave the Way for the Future", *The Hill* [website], 2nd June 2017, available here: https://thehill.com/blogs/pundits-blog/transportation/336125-self-driving-cars-should-help-pay-to-pave-the-way

[215] Marshall, Aarian, and Alex Davies, "Lots of Lobbies and Zero Zombies: How Self-driving Cars Will Reshape Cities", *Wired*, 21st May 2018, available here: https://www.wired.com/story/self-driving-cars-cities/

[216] Harrabin, Roger, "Most New Cars Must Be Electric By 2030, Ministers Told", *BBC News* [website], 17th January 2018, available at: https://www.bbc.com/news/science-environment-42709763

National governments have implemented an array of different measurements and regulation to ensure that safety standards are being met. Many countries have heavily invested in their own independent testing, as there were a number of concerns related to scientific bias in manufacturing testing. In doing so, the US, Canada and Japan have created a greater transparency towards SDV regulation. In total, 65 countries have developed their own SDV driving tests and licensing laws, while also enforcing safety regulations on manufacturers to demonstrate that these vehicles are safe to drive prior to being sold.

There have also been strengthened measures to inform the public about SDVs, how they function, and how non-autonomous drivers should interact with them on the road. This has led to a greater public trust, in conjunction with a large increase in media public awareness campaigns from car manufacturers. There has been a greater emphasis placed on the benefits retrieved from the big data of SDVs, but strict procedures and guidelines have been instituted to ensure personal data is anonymised and encrypted in accordance with GDPR, which has been a milestone for privacy protection over the past seven years.

The automobile industry has had to adapt its earlier approach to the design process of their vehicles, with a greater emphasis on responsible innovation and value-sensitive design. The increase in ethical evaluations of SDVs resulted from state-supported initiatives and the establishment of oversight bodies, such as the UK's Centre for Data Ethics and Innovation, and Singapore's AI Ethics Council. Manufacturers have also had to increase transparency, while also providing guarantees for the life-span of their vehicles. Free software upgrades are mandatory for a five-year period with all SDVs sold in the US, Canada, the EU, the UK, China, South Korea and Japan.

Where software updates occur on a regular basis, manufacturers have provided extensive guidelines about these requirements. SDVs have a built-in locking system that will prohibit drivers from using the cars unless their systems are updated. The cars also have clear and purpose-driven maintenance notification for drivers. Depending upon the seriousness of the maintenance, vehicles may prohibit drivers from operating. There has also been collaboration and agreement through the SDV Fair Use Initiative (SDVFUI) to ensure fair sharing of intellectual property for increasing safety in vehicles.

Since 2023, it has been evident that incorporating more digital infrastructure on our roads would be beneficial for the successful implementation of SDVs. While we are still in early stages, SDVs could be used more optimally with improved digital and physical infrastructure. Civil society organisations have been decrying the possibility that all citizens will have to pay extra for those making the change to autonomous driving, when they are not the ones benefiting from them. Policymakers have been negotiating with SDV manufacturers and owners about paying higher taxes to fund the infrastructure required to accommodate SDVs.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Steps towards a desired future and avoidance of an undesired future

This scenario has covered a lot of ground and outlined many different issues, risks, and possibilities of SDVs in the year 2025. It is very important to reflect on some of these situations and highlight those that are desirable by 2025, those to be avoided, and how to go about doing this. For example, governments should implement appropriate legislation and regulation on the sale, use and safety of SDVs. While national, international and supranational institutions should be responsible for ensuring that citizens are protected from the over-eagerness of manufacturers to put their vehicles on the road. The SDV automotive industry needs to be well regulated and controlled to ensure the safety of the vehicles through the effective implementation of SDV regulatory institutions.

There needs to be adherence to current regulations for the effective control of data generated, retrieved and used by SDVs. Clear delineations need to be established about what constitutes *essential data* for the vehicle's mobility and if this contains personal and private information. There needs to be clear indication that if essential data contains personal or private information, then it should be strongly anonymized, aggregated, and secured, to protect individual's privacy. If it is non-essential data, then there should be adequate policies to ensure that it is not retrieved or stored as a result of using an SDV, unless explicit and informed consent is given. Governments need to effectively integrate the tenets of the GDPR into the automotive industry to effectively assure citizens that their personal data will be protected if they use SDVs. Automobile manufacturers have the responsibility of identifying the purposes for which the car collects data in order to demonstrate their compliance with data protection law. For instance, there needs to be careful analysis if this data will be used for advertising, customised pricing, or to sell additional products to the car owner, and either ensure the owner is aware of these, and consent to it, or prohibit use of data in this way, altogether.

The data collected within the vehicle may become important for law enforcement officials in situations where SDVs are either hacked or being used for malicious purposes. There are already technical options being developed to ensure that the harm caused in these situations is minimised, police authorities identify issues as soon as possible, while at the same time not infringing upon the privacy of innocent citizens using SDVs. Methods such as DENM, certifications, cryptographic signatures, and attestation methods, require heavy investment by automotive companies and need to be fit for purpose. There needs to be careful statutory regulation, third-party testing, and planning for the security of these technologies.

Police need to be granted permission to identify, access, and control vehicles that have been hacked or hijacked. However, they should only

Do you agree with these recommendations?

Are there any others that you think we should include?

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

have access in instances where there is a threat to safety and security, not simply for surveillance purposes. There also needs be effective and appropriate peer-to-peer communications with emergency vehicles, regardless of the fact if they are SDVs or not. Fire brigades, ambulance services, police, or governmental cavalcades, may require access to bypass vehicles and SDVs need to be programmed to identify when these vehicles are approaching. Emergency service vehicles need to be equipped with sensors to inform local SDVs of their approach. One of the undesirable outcomes of SDV implementation in the year 2025 is that the public might have little input into their integration in the market and information about these vehicles. Citizens should be informed about SDV regulation, so it is vital that policymakers receive input and feedback from the public about their needs. Policymakers should consider the needs of all stakeholders, so that policy is created for the public, rather than forced upon them by governments or SDV manufacturers. Policymakers also need to ensure that there is a smooth transition between traditional infrastructure and the digital infrastructure of the future. For the foreseeable future, SDVs will have to use our current road signs, lights and markings to navigate on roads. However, these may eventually be replaced by ‘digital infrastructure’. While this is not likely to transpire by 2025, governments and companies should still begin preparing for this transition. Society needs a sustainable transportation system, and this may either be exacerbated or improved with the proliferation of SDVs. Policymakers need to take careful steps to ensure that they are not ‘overused’, once they become so convenient to use that people start commuting much further from work. One such possibility is greater investment into SDV public transportation systems to ensure convenience, cost and energy reductions. This may also prevent poorer citizens from being excluded from the transportation system. Furthermore, careful attention must be placed on ensuring more inclusionary SDVs, especially when they reach level 4 and 5. In particular, the elderly, handicapped, and those who cannot drive, may be granted accessibility to SDVs.	

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

7. Fifth scenario: Artificial intelligence & robots in education in 2025

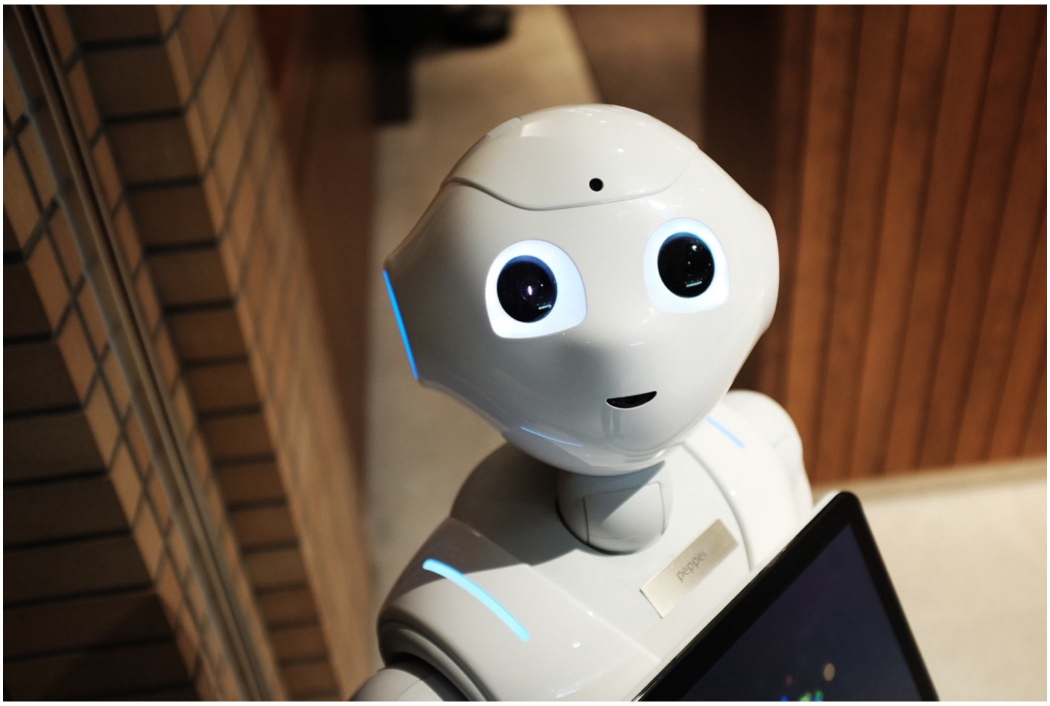


Figure 15 Fifth scenario

1. Introduction

This scenario starts in the frame of 2025 and continues in that frame until the last section where the recommendations suggest measures that policymakers can take now to avoid the undesirable aspects of the future (of 2025) and to reach a desired future.

So off we go to the year 2025, when technological changes have revolutionised the classroom and curricula across schools in Europe. The most important of those changes is undoubtedly artificial intelligence, quickly followed by robots, which have significantly changed education at all levels.

Changes that have been implemented in schools are now moving outside the classroom. Companies are advertising new employment opportunities, tailored to the new curricula. However, it is not clear whether the new curricula and new school structures motivated industry to define new types of workers, or whether the need for new types of workers, a few years back, inspired the push for new curricula and changes in the school environment.

In addition to the industrial push, new government policies are underway to support changes in education (especially in elementary education) in Europe. Their aim is to ensure ethical foundations that can support effective, long-term change.

Currently, the changes in education in the last seven years can be categorised into four types:

1. *Move towards collaborative learning*

F ge
C es

The use of platforms that enable co-creation *between* students has supported group work from a young age, which is now one of the main types of learning activities. Group work supports independent and collaborative learning of skills, as a result of the recent shift towards a skills-centred education paradigm. The platforms are used extensively in teaching the students new, specific skills, e.g., leadership and IT skills, and have been popular in higher education, as they mimic the software development environments of big IT companies.

2. *Use of automation to provide assessment feedback*

Inspired by Jill Watson, a robot used by an individual professor at Georgia Tech a decade ago to provide feedback, robots have taken over the marking of assessments and feedback to students in most school subjects, by using AI-powered software. Similar bots are employed in the service industry.

3. *Personalisation using big data*

Big data collected with the use of the Internet of Things (IoT) has greatly increased personalisation and profiling, a practice especially used for the lifelong learning companions recently launched as an education aid (see below).

4. *Visualisation that allows visiting extraordinary scenarios*

AI-powered software that provides output in terms of virtual reality and augmented reality is used to transform classrooms into gamified learning environments that allow the students to learn by immersing themselves in the environments as though they are part of an electronic game. Gamification in learning has appeared as part of non-formal learning software in the past, but the integration of virtual reality (VR) and augmented reality (AR) versions of the gamification software in the classroom is becoming widespread in European classrooms.

Figure 17 Image Credits: Kingwood freeimages

Education and AI technologies in 2025

Artificial intelligence (AI), as powered and enhanced by big data enabling technologies such as the Internet of Things (IoT), a technology commercialised only within the decade leading up to 2025, is a popular technology, embedded in many products that are used in everyday services and applications.

AI technology is integrated with educational curricula across Europe, such as the robot that supports learning for younger students, known as lifelong learning companion. Whilst the term *lifelong* is currently used loosely, the learning companion joins the young student to support his/her learning experience only during elementary school. In addition, AI is also used for the transformation of classrooms into game-like environments to enhance student experience in schools, and to enhance their skillset.

AI is already a part of everyday life, often without consumers realising that they are dealing with AI-powered software products, not necessarily in the form of humanoid robots, but as AI-powered software integrated in several everyday products. For example, AI-powered bots have replaced service assistants in more than 80 per cent of customer service departments in the retail industry; these bots are using machine learning to respond to customer requests. [217] The availability of huge amounts of data from many different sources has enabled educational technologies based on AI to flourish in 2025.

Educational technologies are redefining the role of the teacher to become a facilitator of the learning activity. This facilitation becomes needed, while simultaneously, demand for the content and information aspects of education, is dropping down. AI-powered assistants and conversational technologies are abundant in 2025. Because information is becoming ubiquitous, teachers employ technology to help students in improving their reasoning and critical thinking skills. Such technology is not intended to provide students with additional information, but helps to identify potential reasoning gaps, suggesting areas and resources that demonstrate different points of view.

[217] E.g. https://singularityhub.com/2017/11/25/8-ways-ai-will-transform-our-cities-by-2030/#sm.0001jjs7fjiokej8s7y1nlx8k73xx, and, https://www.linkedin.com/pulse/ai-economy-2025-eight-trends-shape-our-future-james-canton/

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Nevertheless, the increasing use of AI algorithms and big data does not come without risks relating to gender, class and racial disparities. AI often encompasses bias within its own design and implementation as shown in the following paragraphs regarding a recent product (*classroom robot)* that was biased and had to be withdrawn.

- *A biased classroom robot*

Classroom robots was an AI-powered technology that observed classrooms (as a group of students), and that had to be withdrawn only a year after initial use due to accusations of algorithmic bias (attributed mainly to the design of the learning algorithms, that did not safeguard against the sue of biased training data). These risks arose from the design and implementation supporting the machine learning that took place in the training phase of the AI-based software, and even more so when the software was, in turn, used for teaching support. The classroom robots were meant to observe classroom activity as a whole and identify ways to improve collective learning. However, based on the training data used by these robots the suggestions were identified as biased towards specific content, e.g. if the data set used to draw conclusions from has been collected from a sample that is not free of bias, then the conclusions drawn will not be free of bias, and this will propagate onto the learning process, resulting in a biased AI-product.

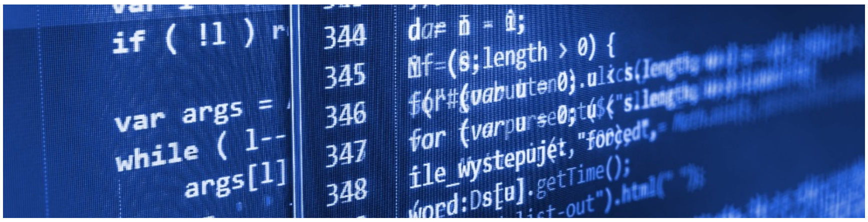


Figure 18 istock, bought by DS

Given the failure of this first attempt at deploying AI in classrooms, guidelines regarding future design of such software were put in place that included aspects such as:

- how to consider bias when selecting training data,
- how to balance transparency against performance issues during implementation, and
- how to plan against bias in data sets that are collected and used dynamically, i.e. without human intervention (e.g. IOT).

Ongoing attempts to create appropriate 'technology in education' related policies across Europe aim to safeguard against similar biases. Although there exist deployments of AI in education in many schools both with student and teacher aids, safeguards against flawed implementations to avoid issues of bias in algorithms are important. Bias mostly arose because of reuse of past AI algorithms and software designs, especially in terms of diversity, an issue only recently improved in the area of AI algorithms.

Applications of AI in education in 2025

Education has not had significant budgetary growth in the past few years, especially as compared to the growth in the technology and business sectors. Yet now, new learning technologies are suddenly available in schools, as predicted by many technologists who have been promoting the integration of technology in education.

To appreciate the significance of these developments, the stalemate that the education sector experienced for a few decades prior to this change must be understood. The World Economic Forum in Davos in 2018 identified budgetary shortages and lack of innovation in the education sector[218], but it still took seven years (2025) to be able to confirm that new learning technologies have been introduced in schools. The World Economic Forum discussed the effects of emerging technologies, especially AI, and the

> "imminent displacement of workers brought on by automation", and added that "there is little doubt that … [education] has severely fallen behind the business world in realising the potential of new technologies – we need to shift our educational mindset to ensure that our children develop skills that can't be replaced by a robot". [219]

Can the use of AI and big data be the tool needed to shift the current educational mindset? In 2025, AI is already present in the education field, through automated assessment feedback and virtual reality and augmented reality spaces. Whilst those have made an impression, they have not managed to motivate new standards and practices in the field, mainly due to the recent change, of introducing AI in schools to support learning, especially in the *lifelong learning companions* for elementary school students.

Schools in Europe have been using a new in-class robot to observe class dynamics and learning trends, in order to identify group characteristics and support the teacher with delivery methods and content. The in-class robot features sociable skills: it has different facial expressions, head positions and tones of voice, which make it similar to a humanoid. Such features help the robot take over the role of teaching assistant.

Teachers are using AI technology in various ways in 2025, one of which is to automate grading of multiple-choice materials as well as more complex types of assessments. For essays or problem-solving assignments, the AI-generated marks are often matched to a human assessment marker as a verification tool.

Some university classes have hundreds of students. AI-powered robots help overworked professors to answer thousands of questions over the course of the semester.[220] These robots can answer any curriculum-

[218] (Baker, 2018)
[219] (Baker, 2018)
[220] Ashok Goel, an academic at Georgia Tech, has described how an AI-powered robot helped him (Goel, 2009).

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

related question over an online classroom space without the students ever realising that they were talking with a robot. Similar robots are used in online classrooms to provide teaching support and student feedback to thousands of online students at a time. Student feedback verifies that the experience is not different from interacting with the class professor over the virtual classroom space.

Some students in 2025 have the opportunity to use an AI lifelong learning companion, as the learning buddies are entering many European classrooms. The AI-powered learning companions adapt to each student's individual strengths and weaknesses in an effort to provide learning assistance throughout their student's life. [221] AI-powered learning companions help students with special needs by adapting materials to lead them to success, as well as providing personalised tutoring for students outside of the classroom. When students need to reinforce skills or master ideas before an assessment, the AI-powered learning companion provides students with the additional tools they need, like revision lists based on their personal learning style or study guides organised according to the students' preferences. Classroom robots learn to identify classroom weaknesses, such as when groups of students miss certain questions, and inform the teacher when material needs to be retaught. In this way, AI can also hold teachers accountable and strengthen best teaching practices. Nowadays, in 2025, these AI learning companions are a reality in many European countries, trained often by observing the young students and conversing with them. The more data the robot collects over time, the more accurate the student profile becomes, and the more helpful the robot's support.

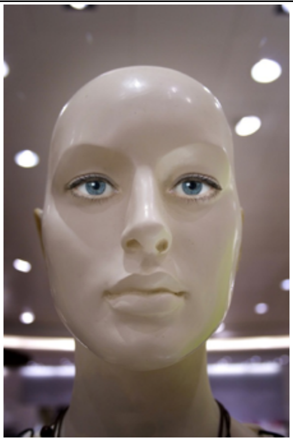


Figure 19 Image Credits: Jean Scheijen freeimages

The classroom as a concept is becoming redefined. This is a result of learning technologies making the classroom boundaries permeable to the learning content, but also because these technologies are becoming increasingly social, helping students to form learning communities based on their interests and skills. The grading system has also spanned outside of the classroom to the subject- and domain-specific learning communities. Teachers who develop content for such learning communities start to collaborate across national borders. Basic community courseware already utilize AI-powered facilitation, while advanced courses are requiring human facilitators to intervene online and offline.

2. Vignette

[221] https://www.pearson.com/corporate/about-pearson/innovation/smarter-digital-tools/intelligence-unleashed.html

Robots in classrooms, often referred to as learning buddies, are now a reality in many European countries. Learning buddies are AI-powered robots that support individual students' learning experiences by building their educational profile over time and using input data from the students themselves. The learning buddies learn by observing the young students and *conversing* with them, and, the more data the robot collects over time through these interactions, the more accurate the student profile becomes. Educationalists now believe that by allowing the learning buddy to accompany students throughout their educational experience, including home life, it can provide more accurate and constructive support to enhance the student's learning experience. Learning buddies are a European initiative that has not been welcomed enthusiastically by all Member States. Issues of cost and training, in addition to the long time it requires to have a quantifiable impact on education, have led many European countries to reject the initiative, even though others have welcomed the initiative with enthusiasm. In the 2025 academic year, for the first time, students can take the learning buddies home with them, and keep them throughout their elementary school education. The artificial educators are expected to have a positive impact not just on students' learning and studying habits but also on their overall social presence, as there is a particular focus on skills such as decision-making, rather than on particular content or information. The development of learning buddies as a mass market was an exemplary collaboration between computer scientists, educationalists and roboticists to work together more closely to support this innovative educational approach. Given the interdisciplinary demands of this initiative, policymakers have begun officially to encourage such collaborations, to avoid another unsuccessful attempt at incorporating AI in education, as happened before with biased classroom robots (see above). The new learning buddies have been more successful as they support transparency of decision-making and training data, as well as open source code. The integration of the learning buddies in the students' home and family life is a new challenge that needs to be collaboratively addressed by education professionals and software developers as well as parents and students. Many open questions remain, including these: • How are different attitudes across Europe affecting the issue of AI in education? • How does the digital divide across Europe affect this opportunity? • Can we apply best practices uniformly after determining the level of success of this technology? • Will this development positively or negatively affect wellbeing of students and teachers overall? • Can this effect be quantified?	Do you think the use of a vignette helps to makes it easier for stakeholders to relate to the technology and its impacts?

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

- Who is accountable for data training?
- How will the algorithms used be chosen?
- How do policymakers plan against bias in developing and using learning buddies, against discrimination in terms of learning, and against inequality in terms of access and liability upon usage?

3. Drivers

There are many actors and stakeholders transforming the education field. The primary driver for achieving a technological change in education that can power rapid growth of the field as envisioned by educators is the robotics industry which sees a vast market for learning buddies.

In addition, educators have long recognised the need to upgrade teaching methods, content curriculum change and to take advantage of new technologies.

Another significant driver is the personal need for the well-being and happiness of students and teachers, who often identify the experience of interaction with state-of-the-art technologies as happy in part because of the playfulness designed into the robots ("learning is fun"). The introduction of learning buddies has allowed teachers to focus on core educational tasks instead of tedious administrative tasks.

Social drivers also *play* a key role, such as the need for diversity. Government policies favourable to the introduction of learning buddies have also played a role.

Market trends

Market trends play a significant role in the deployment of the specific technology in education settings. The degree of autonomy and financial room that schools have, in order to adopt new technologies plays a significant role in how these trends are playing out. Budget assignment in public education is always challenging, but the trend of AI use in education is popular in the private sector. Industry has been sponsoring educational projects for years, because they need more data scientists, roboticists and other high-tech professions.

The advance of technology continues to create new job opportunities in the industrial sector while the education sector has become more attuned to the need to supply candidates with the skills that industry demands. Industry has given the education sector a push to encourage the sector to collaborate with industry in supporting emerging technologies. There is huge competition around the world for data scientists, in particular. The education sector and politicians have

Do you agree with these drivers? Are there any other significant drivers that should be included here?

recognised Europe’s need to compete with the US and China, where data scientists command extremely high salaries as Google, Apple and others have attracted most of the talent, even before students graduate from university.

Need for renewal of teaching, learning content and methods

Technology in education provides an opportunity for change and innovation in terms of the teaching and learning approaches as well as in the content and the teaching mode itself, in addition to using technology for curriculum delivery.

Technology is affecting the content of what is taught in schools and how it is taught. This has resulted in new profitable opportunities in the education sector, e.g., for lifelong learning support and non-formal education. Moreover, teaching methods have changed by 2025, as the teaching and learning environments incorporate technology and much of the curriculum content is available online, with students having easy and frequent access. The need for change and enhancement of teaching and learning content and methods has been necessary and, at times, difficult. Some experts have considered AI-powered software to support this change as the easy option compared to the alternative of restructuring schools and curricula or re-educating the teachers themselves.

Within the context of teaching and learning changes, there has been a need for new types of classrooms, more scalable solutions and flexibility in the timing of the learning itself. Technology offers a perceived flexibility in learning modes, which the learning environments support. The emerging technologies have helped create a multimodal learning educational system that promotes relief from the traditional information overload. Multimodal learning offers information in multiple formats, in addition to the traditional classroom teaching. Hence, in 2025, the education system has become a hybrid with traditional school components but also another component outside the classroom as a support for diverse learners.

Need to support curriculum change

In 2025, ministries of education have developed curricula focused on skills, especially critical reflection skills rather than information. AI-powered software provides the information, and hence students focus on enhancing skills, rather than memorising information. Ministries of education

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

recognise the importance of character-building curricula aligned with the identification of critical and innovative thinking[222].

Need for personal satisfaction

Students and teachers need to be happy and satisfied, a psychological driver that supports learning and is key in research within the area of affective computing. Technology is arguably a popular method to increase students' interest and engagement, whereas teachers welcome any support that will relieve them from some of their tasks, e.g., keeping track of progress and content response. Expanding the information base to which students have access through the AI-powered software supports diverse cultural backgrounds and provides a common platform for learning and encouraging diversity and social connections, as an enhancement of peer-to-peer socialisation. This improves overall student experience. Classrooms and schools have been upgraded by the AI-powered teaching aids, such as augmented reality software for teaching and learning support.

The technology has helped stabilise or even reduce overall student expenditure, as parents no longer need to invest in personal tutors, although this may be an issue that varies among parents of differential financial capabilities, as the leaning buddies support students' learning inside and outside the classroom.

Social driver: Support for learning diversity

The decreasing size and increasing usability of new technology have turned learning into an agreeable experience for young students. Technology hype is a factor in offering diverse learning opportunities outside the classroom. Current trends of integrating AI in education are growing as the actual technology has become more robust. Before the widespread adoption of AI in education, most educators recognised that the traditional education system was obsolete. AI has created revolutions in all sectors of our society, economy and policy, including the education sector. Diversity and the perceived freedom from educational bias are attractive to learners, further reinforced by a public sentiment of AI technologies as hype and fashion. In 2025, all EU countries have national ethics committees who require developers to ensure such technologies are free from bias and address related ethical risks.

[222] (UNESCO, 2015)

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Technology driver: Emerging technologies, adaptive learning technologies

Another driver for the huge increase in the use of AI in education is the flexibility, adaptiveness and computational power of the new technologies achieving feats of power that even experts did not predict a mere four or five years ago. Universities have established centres of excellence focused on adaptive learning technologies and studies of the computer-brain interface and interactions.

In 2025, learning tools are personalised, mainly because of adaptive learning technologies that adapt to the education needs of students. AI gives students real-time feedback to ongoing lessons. In 2025, students know more, understand more, process information more efficiently than they did a mere seven or eight years ago. Moreover, today, students are using cognitive enhancements to achieve accelerated learning. Such enhancements are unevenly distributed across the population, partly because of cost and partly based on the sheer availability of the technologies and specialised apps, including educational games, new types of learning spaces, virtual and augmented reality, etc.

Political driver: New government policies

The increasing public awareness of these technologies and the issues they raise regarding privacy, security and ethics have led to the formation of new grassroots groups. These are putting pressure on policymakers to address these issues in a responsible way, in the public interest. The General Data Protection Regulation (GDPR) helps but does not address all of the issues arising from the use of AI in education – such as the uneven distribution of these technologies across the population. Accessibility, bias, discrimination, the varying needs of different socio-economic groups and policy enforcement are all issues that beset educational ministries in 2025. Citizens agitate for new government education policies addressing issues of trust and distrust, ownership of data, learner analytics, cost-cutting, and new alternative learning methods, including the use of AI in education.

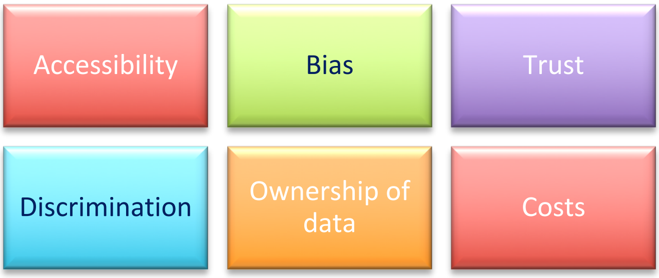


Figure 20 Concerns for Governments

4. Barriers and inhibitors

The most significant barriers and inhibitors in 2025 that can hinder the integration of AI in the education field, include most likely the constraints outlined below, such as economic constraints, social constraints, resistance to change by the educators themselves and for curriculum change as well, and finally barriers that exist by technology itself, e.g. technology products that collect and interpret data in a biased manner, due to errors in software and data collection design.

Fears of unemployment

Teachers, administrators and others in the sector are apprehensive about their future employment, as AI and robots are increasingly used across the sector. Learning buddies are still a status symbol. Some schools offer them, but not all do so. However, the widespread adoption of the technology by most schools is driving down the price of learning buddies to levels affordable by most (though not all) middle-class families in Europe. Despite such fears, the adoption of learning buddies is creating new employment opportunities for constructing, marketing and servicing the robots.

Regulatory issues

The educational content conveyed by the learning buddies has become an issue. Who decides the learning content of the learning buddies? This issue refers mostly to having policies regulating the content of the training datasets, which provide checks for possible bias in the learning process of the learning buddies.

Economic constraints

Optimising educational opportunities is not only a matter of the availability of technology, but also the creation of environments within schools to support this technology. The AI-powered education systems have not been cheap, but their cost is falling rapidly. One of the cost factors has been the need to further investigate issues of bias and security during its deployment in schools. Industry has, of course, given a vote of support for the policy changes in the education sector, as it increases their penetration of these markets for AI-based solutions. For relevant, budget-related government policies for education, especially in public schools, AI supported learning has emerged as a governmental priority, as it works towards addressing issues such as trust and distrust, ownership of data, learner analytics, effectiveness of alternative learning methods, and much else. These issues are positive drivers with potential to become inhibitors.

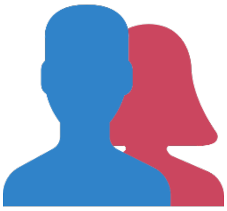
Do you agree that these are likely to be the most significant barriers and inhibitors in 2025?
Are there any other barriers that should be mentioned?

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

Social acceptance

Learning buddies have become status symbols. Some schools offer them, others don't. Those that do are generally in more privileged areas. There are issues of social envy, discrimination, unfairness, class prejudice, inequality as some feel that offering these advanced services and technologies to some schools, not all, is discriminatory. Scepticism has also been an inhibitor to the wide use of such technologies, because of the fear of some school boards of something new. The public is conflicted on the use of robots. Some people fear that big data means intrusions upon their privacy. Some fear privacy invasions, risk of manipulation, and of targeted advertising. The fear of pervasive surveillance has, ironically, escalated to fear for safety and security of their sensitive personal data in so many different places, some of which come together in the hands of data brokers. There are few public consultations on the introduction of AI and robots in various public sector departments.

Resistance to change: educators



The one-size-fits-all approach of school boards and education ministries was shed in favour of a more adaptive, inclusive, open approach toward formulating curricula, taking into account the demand for particular jobs over the next decade. These changes required both physical changes to the school infrastructure and environment and content changes to the curriculum itself. Traditionally, changes to the curriculum and approaches to teaching have been time-consuming and costly. The resistance of some teachers to the use of technologies because they fear those technologies could lead to their unemployment closely parallels overall social acceptance too. This factor relates closely to the inhibitor of social acceptance, especially the degree to which educators themselves are willing to support a more widespread use of robots and other AI system or whether their fear of being replaced will act as a deterrent to the adoption of new and emerging technologies and especially their integration into comprehensive networks.

Resistance to change: curriculum

As the need for change of curricula has been a key driver for the better functioning, more optimised education system, the scale of the challenge of achieving this change has been a daunting inhibitor. Change of curricula is a time-consuming process according to established educational policies and methodologies. In addition to time, the decision-making of the change is often difficult,

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

i.e., who will propose the new learning goals and how can it be ensured that these new learning goals are free from bias and ethical concerns? Changing the curriculum in order to support new, technologically-enhanced learning goals is different across Europe. This concerns not only the level of digital literacy in each country but also local policy, available budget, etc. Nevertheless, empirical evidence based on a selection of assessment criteria, from the trials in 2023-2024, have shown that schools with class robots and learning buddies outperformed schools without such technologies. They also found a high acceptance rate among students. *Technology barriers* By 2025, companies and research organisations are producing technology based on privacy by design, ethics by design and security by design. An inhibitor towards this direction is the limitation of technology itself. The technology needs vast amounts of data to support unbiased learning, and such data is not always available.	
5. Ethical, legal, social and economic impacts As AI becomes more powerful, more autonomous and broader in its use and impact, unresolved ethical implications of AI are a challenge, with unpredictable developments, involving ethical and social risks such as discrimination, inequality, unfairness, bias, lack of transparency, job losses, privacy breaches and malevolent use of the tools themselves. Educationalists, policy-makers, teacher-parent associations and other stakeholders have constructed a framework that governs the development and use of these technologies in an educational environment where students interact repeatedly with these tools, exposing personal data, habits, learning styles and, moreover, are expected to trust the AI-powered learning aids. Unavoidably, data must be shared to allow the AI-powered learning support applications to make better decisions and help students and classrooms. Therein comes the challenge that all data has to be kept safe and anonymous. As risks are more clearly identified, better solutions come into focus.	*Ethical issues* Do you agree that the ethical, legal, social and economic impacts listed here are likely to be important in 2025? Are there any other issues to which we should refer?

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

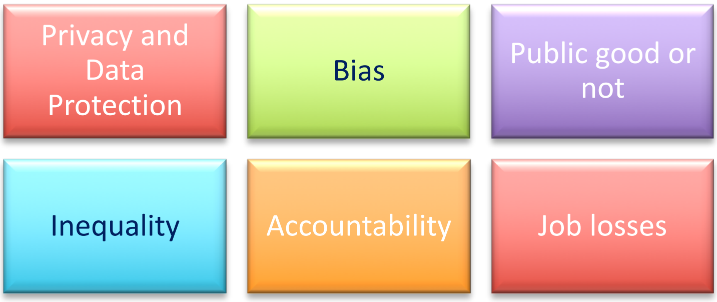


Figure 21 Ethical challenges

Privacy and data protection

The central ethical issue identified is the issue of privacy, since there will be a generation, collection and manipulation of personal data, specifically of sensitive personal data. The class room robots and learning buddies are constantly collecting data from their environment, including interacting with the students, via video and audio monitoring (surveillance) of the human.

Although the GDPR has curtailed unlawful use of personal data, vast amounts of personal data are collected during the education-enhancement process that must be regularly investigated and regulated. For instance, what will be recorded exactly and who will have access to this data? Will this specific education- supporting data include student competencies and student emotions?

Bias

Bias continues to be an issue in 2025. Bias is a risk because of AI's learning capabilities. It can learn bad things as easily as good ones, since AI learning is based on training datasets and unless these are carefully generated, there exists at least the risk of collecting data that is not representative of an unbiased population sample. Bias appears during the different stages of AI development: while framing the problem, while collecting the data and while preparing the data to be used by machines. It is hard to fix because of unknowns, imperfect processes of data collection and annotation, lack of social context and most importantly, different definitions of concepts such as fairness.

Public good or not

Industry views schools as a service, as a resource for industry's recruitment needs and as a new avenue for business, instead of viewing education as a public good.
Accessibility and inclusiveness remain important social issues in 2025, with sufficiently wide public support that industry ignores at its peril.

Inequality and asymmetries

Ownership and access to the technology contribute to inequality (in terms of opportunities) and to information and power asymmetries, often involuntary. This can be the result of AI-skilled humans or even different capabilities of human-to-robot interaction.

Accountability

Moral responsibility and accountability are still an issue, highlighting possible implications for students, teachers, parents, school administration, policymakers and software developers.

Freedom of thought

There is a further concern that by introducing frequent human-robot interactions, the human free thought will be affected by the robot decision making process. Questions on how to ensure that technology does not affect freedom of thought in students, especially easily impressionable young students, still arise. Such issues must still be addressed by software developers and checked by policymakers. Furthermore, the freedom to make mistakes to go through their own learning process and growth is a concern that relates to the issue of hindering freedom of thought.

6. Recommendations for a desired future and avoiding an undesired future

Considering the current identification of possible inhibitors to the design and use of AI in education in 2025, there is the opportunity to mitigate these negative aspects and accentuate the positive aspects for the introduction of AI in education in 2025.

The introduction of AI in education can have long-term potential, especially once the specific AI algorithms employ a new design approach that secures the generated software will be as free of bias as necessary. It is also important to safeguard against dependencies that social intelligence may create, e.g., students' attachment to AI-powered learning buddies.

When designing AI-powered systems for education, conclusions or final decisions should not be made by the systems, even though the systems that support AI can also make intermediary decisions. Educators should protect free thought and support students' skills' enhancement so that they make informed final decisions. Achieving this will need an understanding of the educational landscape across Europe and how it

Do you agree with these recommendations? Are there any others that you think we should include?

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641

can easily support accessible and acceptable changes to mitigate scepticism and accentuate growth.

Next, we list and describe selected recommendations related both to governance and ethics as well as to education and learning.

AI governance and ethics

Human control of AI decisions

A first recommendation is to maintain human control over the use of AI, in an attempt to eliminate many of the problems associated with fully autonomous systems. Such a requirement would protect the dignity of human life, freedom of choice, facilitate compliance with international humanitarian and human rights laws, and would promote accountability for unlawful acts. System design may promote suggestive or verification support provided by the AI systems.

Transparent use of affective computing

Machines use affective computing and AI techniques to sense, understand, learn and interact with human emotions. A combination of facial recognition, gait, language, voice pattern analysis, can already decode human emotions with a high degree of confidence. Acknowledging human emotions, using them as factors for the decision-making, as well as influencing emotions to bring the value to the AI solutions should be transparent. A higher degree of openness could be achieved, for example, by using Open Ethics Vector[223] to transparently communicate algorithmic approaches.

Value alignment for AI systems
Considering that human behaviour does not always reflect human values, then AI systems, even though they are able to learn a lot by observing students and teachers, may be fundamentally unable to distinguish between value-aligned and misaligned human behaviour to provide AI educational products with appropriate learning feedback. A recommendation towards addressing such inconsistencies, is to make use of a value-alignment mechanism to help system distinguish between value-aligned and misaligned human behaviour.

Dealing with bias

Eliminating intrinsic bias caused by training approaches is a crucial step towards making AI systems effective learning aids. Students should be brought to an awareness about the processes and events which are caught by the AI's "attention". Systems should be providing students and educators with the ways to incorporate additional information as well as to make final choices and decisions and the degree to which AI systems influence human decision should be explicit. Moreover, the

[223] https://openethics.ai/

communicational distance should be kept so that the system operates in a non-manipulative manner. **Education design and future classroom** *Create facilitation environments and promote inquiry-based learning* To reach the desired future, the design and implementation of new educational environments must be considered. Creation of facilitation environments and subsequent transformation of current one-to-many teaching model of the classroom, is recommended. Facilitation environments focus on students achieving their learning goals by using project-based learning. The freedom that this approach gives students for selecting their own learning pace and style, allows them to enhance their independent study skills by receiving help on specific matters based on their performance and questions. Teachers should be encouraging divergent thinking and allowing students the freedom to ask their own questions and to learn the effective strategies for discovering the answers. *Focus on student coaching* Adopting a process to help students in their professional orientation by helping students to learn more about themselves using introspection approaches, discovery sessions, as well as by adopting coaching technology in and outside of the classroom, is recommended. *Work with real-world problems and data* Students should work with real-world data in their school assignments. The open innovation platforms and questions could be supplied by the public and private organizations. The degree of complexities for problems supplied by such open innovation system could be evaluated and assigned to students by educators. The problems could span from simple data annotation, collaborative evaluation of each-other’s work to crowdsourced solution of challenges for urban planning and local economy.	

8. Conclusions and recommendations

8.1 The SHERPA scenario construction process

A novelty of Task 1.2 was our scenario construction process, involving stakeholders from the outset and throughout the process. The process was structured, starting with the workshop participants and expanding to the project’s stakeholder board, its contact list and the public at large.

The scenario workshops were also structured. The workshop agendas tracked the structure of the scenarios themselves. The structure we created for the scenarios was used in all five instances, with only slight variations.

The scenario construction process has been a useful opportunity to engage with a diverse group of stakeholders, to share views about how smart information systems, notably artificial intelligence, are changing and might further change our society and economy. In all five scenario workshops, the discussion never stalled; participants positively bubbled over with ideas and views about how AI might develop by 2025 and the impacts those developments might have, even though most of the participants had never met before. So, scenario brainstorming worked in getting all participants to express their views and to participate.

8.2 Key conclusions

One of the objectives of the scenario construction process was to reach a consensus on a **plausible future**. Participants were challenged to be creative, to leap ahead six or seven years and imagine how the technologies might evolve and what new applications might arise. The present often got in the way of the future in many of the discussions, but mostly the present provided a reality check on a story-telling exercise. We settled on the year 2025 – seven years away at the time of the workshops – as not being so far away that we needed to enter the realm of science fiction, nor being too close in term of simply building on what exists today. We wanted more than the present, as it were, without getting trapped in science fiction. We sought to develop plausible scenarios with **recommendations** that would be useful **for policymakers**. The scenario construction process, as in SHERPA, is a way for policymakers to get "ahead of the curve", to develop policies now that will anticipate or pre-empt an undesired future and promote a desired future. In other words, the policy development process needs to begin now, as it usually takes several years before an identified policy requirement becomes legislation.

Two of the workshops benefited from the presence of an EC policy officer, who urged us to develop recommendations that were practical and, preferably, coherent with the recommendations from two other EU-funded projects, SIENNA and PANELFIT. The co-ordinators of the three projects have been in contact and agreed to collaborate, e.g., in invitations to workshops, sharing deliverables, etc. Two of the partners in SHERPA are also the co-ordinator and deputy co-ordinator of the SIENNA project.

It is already obvious that artificial intelligence is having far-reaching impacts and those impacts will only amplify as the technology evolves, as algorithms are used in ever more applications. Some applications, e.g., Google Translate or the Duck Duck Go search engine, are useful, while others (such as targeted advertising) offend many consumers and invade their privacy and misuse people's personal data without their explicit and informed consent. Hence, participants saw AI and other smart information systems as bringing **great benefits, but also great threats**.

While some of the recommendations from the five scenarios are specific to a specific technology area, there are some common themes that appear – and **data protection** is one of those. Workshop participants were concerned about the use of personal data without the consumer citizen's consent. They were concerned about AI being used to improve advertising targeted at individuals and other invasions of **privacy**.

Another common theme was the need for greater, more coherent **regulatory oversight** in the application of the technologies. Participants all agreed that artificial intelligence, while offering great benefits, also creates great risks. Sometimes malefactors deliberately develop AI systems and algorithms to achieve their gains at the expense and harm of society. This is what happened in the WannaCry take-down of the UK National Health Service, which seems to have been a clumsy, but moderately successful plot by North Korea to earn some foreign revenues at the expense of the NHS and various other organisations.

Gigantism was another common issue – although that specific word was not used, still there was a shared sense that that the big five companies – Amazon, Apple, Facebook, Google and Microsoft – wield far **too much power** with little effective oversight. The big five, in effect, control the AI market. They hoover up much of the AI talent. Their resources and the amount of data at their disposal dwarfs anything by any other player. Hence, the big five are driving the future of AI and putting algorithms to work in a vast array of different applications to understand us better, in a phenomenon that Zuboff calls "surveillance capitalism".[224]

There was some discussion about the need to bring **explainability** into algorithms – i.e., algorithms had to inform users or those affected by the algorithms the purpose, who was funding the development of the algorithm, whom to contact for more information. This rarely happens now but participants hoped it would more likely be the case by 2025.

Another issue that arose was that of **inequality**, i.e., that some people were more likely to benefit from AI (e.g., from robotic learning buddies or holographic companions) than those in a lower socio-economic stratum. The related issue of **fairness** also arose, e.g., predictive policing algorithms were more likely to target street crime than corporate crime.

As AI penetrates further into our economies and societies, it is **speeding up decision-making** such that AI-powered decision-making becomes more needed. Human decision-makers cannot respond fast enough, especially in the instance of attacks on cities and critical infrastructure. AI-powered decision-making raises apprehensions about decisions gone wrong or without an appreciation of the consequences.

AI often raises **complex ethical** issues, especially regarding legal and moral liability. Some AI scientists have already signed a petition against working on killer robots; some employees have rebelled against working on AI military technologies. Questions of **liability** proliferate. Who is liable for an algorithm on which many data scientists have worked? Is it the organisation who is funding development of the algorithm? Is it the programmer who feeds the data to train the algorithms? Is it the client who is using the algorithm? Do the middlemen, the suppliers, have some liability? Or the insurance companies? Other issues worth debating are those relating to **autonomy**. Is AI creating dependencies, and thereby reducing our autonomy? Some of these issues are also being explored in the SIENNA project.

8.3 The bottom line - recommendations

It is obvious from the scenario-construction process and from the scenarios themselves that AI offers many benefits and raises many threats and ethical issues, especially in regard to third-party unauthorised use of personal data, intrusions upon our privacy, manipulation of social media, consumers and citizens, and ready-made opinions. AI pervades our societies and economies and will increasingly do so. It will affect us as individuals and collectively as communities and societies. The transformational power of AI far exceeds other regulated products and services, such as cigarettes, highways, medicines or industrial waste, yet AI goes largely unregulated or only partly regulated in some narrow areas. For all intents and purposes, AI goes unregulated.

None of the scenarios discussed regulatory models or went into any depth on the nature of appropriate regulatory models, but all reflect the need for some form of regulation. The diversity of issues and applications illustrated by the scenarios suggest that regulation needs to be multidisciplinary in scope. One of the recommendations in the first scenario stated: "Existing regulators should adopt a co-ordinated (co-regulatory) approach to AI mimicry to ensure harmonised, consistent rules for industry. As holograms like

[224] Zuboff, Shoshana, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, Profile Books, 2019.

Lucy raise various issues beyond the remit of a single regulator, some mechanism is needed to ensure regulatory harmonisation."

Most regulators are sector specific[225], but AI crosses all sectors. To be effective, a regulator needs enforcement powers. A new regulator with a remit to challenge AI practices in whatever domain may lead to conflict with sector-specific regulators. So, when policymakers and legislators are thinking about regulatory options, they will need to take into account the sensitivities and the mandates of other regulators (where they exist).

Regulatory options are the subject of future SHERPA deliverables, but suffice it to say here, based on the scenarios and as an input to those later deliverables, that any new regulator or regulatory scheme will need to consider the inclusion of a wide range of competencies – technical, legal, ethical, organisational, economic, political, cultural – with enforcement powers across sectors and jurisdictions and with the sensitivities and diplomatic skills required to interact with other regulators, some of whom will already have formidable powers of their own.[226]

Furthermore, AI-powered technologies cross borders. The scenarios do not suggest situations confined to specific countries. Hence, any regulatory scheme will need to have a trans-border, international dimension. Furthermore, as the scenarios depict, AI touches the lives of many (even most) consumers and citizens, hence, any new regulatory scheme will need to raise public awareness about the dangers of AI. The benefits speak for themselves, but the dangers hide in black boxes.

[225] The US Federal Trade Commission is an example of a regulator with powers that extend across many sectors in the economy.

[226] Interestingly, a few days after we wrote this comment, the House of Lords called for a super-regulator. See Hern, Alex, "House of Lords report calls for digital super-regulator", *The Guardian*, 9 Mar 2019: "The House of Lords has called for the creation of a digital super-regulator to oversee the different bodies charged with safeguarding the internet and replace the "clearly failing" system of self-regulation by big technology companies. A new Digital Authority is the chief recommendation of the Lords' communications committee report, which warns that the patchwork quilt of more than a dozen regulators that oversee the digital realm creates gaps and overlaps." The chair of the committee, Lord Gilbert of Panteg, said, "Self-regulation by online platforms is clearly failing and the current regulatory framework is out of date. The evidence we heard made a compelling and urgent case for a new approach to regulation. Without intervention, the largest tech companies are likely to gain ever more control of technologies which extract personal data and make decisions affecting people's lives."

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement no. 786641