National program for promoting the development and use of artificial intelligence in the Republic of Slovenia by 2025 (NpUI)

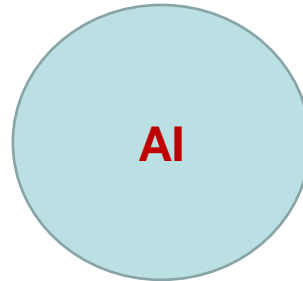# -

how to handle bias?

Samo Zorc, Coordinator of preparation of NpUI

Ministry of Digital Transformation

16.5.2022, EU AI4GOV workshop on bias

1. **AI evolution**
2. **NpUI – AI system definition + bias**
3. **NpUI – concept, context, strategic objectives**
4. **AI Act – handling bias for HR AI**

THE REPUBLIC OF SLOVENIA
MINISTRY OF DIGITAL TRANSFORMATION

AI

THE REPUBLIC OF SLOVENIA
MINISTRY OF DIGITAL TRANSFORMATION

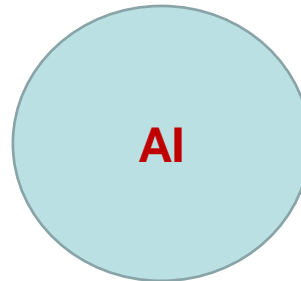ITU Events https://aiforgood.itu.int/event/ai-powered-vehicles-for-humanitarian-help-deployment/

AI for Good Webinars

AI-powered vehicles for humanitarian help deployment

Thursday, 31 March 2022
16:00 - 18:00 Geneva (CEST)

In cooperation with
INNOVATION ACCELERATOR

aiforgood.itu.int

WFP World Food Programme

AI

THE REPUBLIC OF SLOVENIA
MINISTRY OF DIGITAL TRANSFORMATION

ITU Events https://aiforgood.itu.int/event/ai-powered-vehicles-for-humanitarian-help-deployment/

AI for Good Webinars

AI-powered vehicles for humanitarian help deployment

Thursday, 31 March 2022
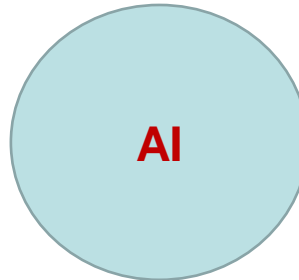16:00 - 18:00 Geneva (CEST)

In cooperation with
WFP World Food Programme

aiforgood.itu.int
INNOVATION ACCELERATOR

Solving problems in medicine
https://www.nature.com/articles/s41598-019-51147-3.pdf

Our data demonstrate the feasibility of brain tumour diagnosis from routine blood tests using machine learning. the reported diagnostic accuracy is comparable and possibly complementary to that of imaging studies. the presented machine learning approach opens a completely new avenue in the diagnosis of these grave neurological diseases and demonstrates the utility of valuable information obtained from routine blood tests.

AI

THE REPUBLIC OF SLOVENIA
MINISTRY OF DIGITAL TRANSFORMATION

ITU Events https://aiforgood.itu.int/event/ai-powered-vehicles-for-humanitarian-help-deployment/

AI for Good Webinars

AI-powered vehicles for humanitarian help deployment

Thursday, 31 March 2022
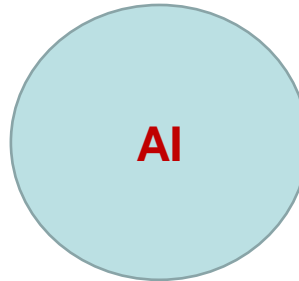16:00 - 18:00 Geneva (CEST)

In cooperation with

World Food Programme

aiforgood.itu.int

INNOVATION ACCELERATOR

Solving problems in medicine
https://www.nature.com/articles/s41598-019-51147-3.pdf

Our data demonstrate the feasibility of brain tumour diagnosis from routine blood tests using machine learning. the reported diagnostic accuracy is comparable and possibly complementary to that of imaging studies. the presented machine learning approach opens a completely new avenue in the diagnosis of these grave neurological diseases and demonstrates the utility of valuable information obtained from routine blood tests.

Solving environment problems

Plastic pollution of see
https://www.sciencedirect.com/science/article/pii/S0269749121000683?via%3Dihub

AI

THE REPUBLIC OF SLOVENIA
MINISTRY OF DIGITAL TRANSFORMATION

ITU Events https://aiforgood.itu.int/event/ai-powered-vehicles-for-humanitarian-help-deployment/

AI for Good Webinars

AI-powered vehicles for humanitarian help deployment

Thursday, 31 March 2022
16:00 - 18:00 Geneva (CEST)

In cooperation with

World Food Programme

aiforgood.itu.int

INNOVATION ACCELERATOR

Solving problems in medicine
https://www.nature.com/articles/s41598-019-51147-3.pdf

Our data demonstrate the feasibility of brain tumour diagnosis from routine blood tests using machine learning. the reported diagnostic accuracy is comparable and possibly complementary to that of imaging studies. the presented machine learning approach opens a completely new avenue in the diagnosis of these grave neurological diseases and demonstrates the utility of valuable information obtained from routine blood tests.

AI

Solving environment problems

Plastic pollution of see
https://www.sciencedirect.com/science/article/pii/S0269749121000683?via%3Dihub

Prediction of floods in coastal reagions
https://doi.org/10.3986/alternator.2021.35

AI Omnipresence – Physical World

ITU Events https://aiforgood.itu.int/event/ai-powered-vehicles-for-humanitarian-help-deployment/

AI for Good Webinars

AI-powered vehicles for humanitarian help deployment

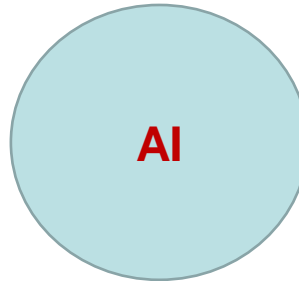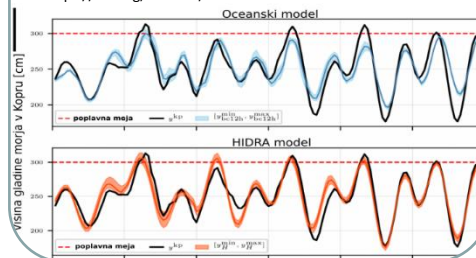Thursday, 31 March 2022
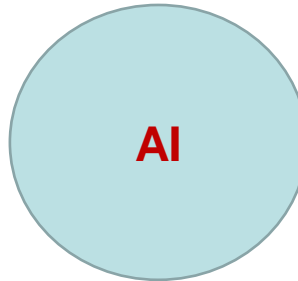16:00 - 18:00 Geneva (CEST)

In cooperation with
World Food Programme

aiforgood.itu.int

INNOVATION ACCELERATOR

Solving problems in medicine
https://www.nature.com/articles/s41598-019-51147-3.pdf

Our data demonstrate the feasibility of brain tumour diagnosis from routine blood tests using machine learning. the reported diagnostic accuracy is comparable and possibly complementary to that of imaging studies. the presented machine learning approach opens a completely new avenue in the diagnosis of these grave neurological diseases and demonstrates the utility of valuable information obtained from routine blood tests.

AI

Solving environment problems

Plastic pollution of see
https://www.sciencedirect.com/science/article/pii/S026974 9121000683?via%3Dihub

Prediction of floods in coastal reagions
https://doi.org/10.3986/alternator.2021.35

Predictive maintenance

https://en.wikipedia.org/wiki/Predictive_maintenance

Predictive maintenance is a proactive maintenance technique that uses real-time asset data (collected through sensors), historical performance data, and advanced analytics to forecast when asset failure will occur.

THE REPUBLIC OF SLOVENIA
MINISTRY OF DIGITAL TRANSFORMATION

ITU Events https://aiforgood.itu.int/event/ai-powered-vehicles-for-humanitarian-help-deployment/

AI for Good Webinars

AI-powered vehicles for humanitarian help deployment

Thursday, 31 March 2022
16:00 - 18:00 Geneva (CEST)

In cooperation with
World Food Programme

INNOVATION ACCELERATOR
aiforgood.itu.int

Solving problems in medicine
https://www.nature.com/articles/s41598-019-51147-3.pdf

Our data demonstrate the feasibility of brain tumour diagnosis from routine blood tests using machine learning. the reported diagnostic accuracy is comparable and possibly complementary to that of imaging studies. the presented machine learning approach opens a completely new avenue in the diagnosis of these grave neurological diseases and demonstrates the utility of valuable information obtained from routine blood tests.
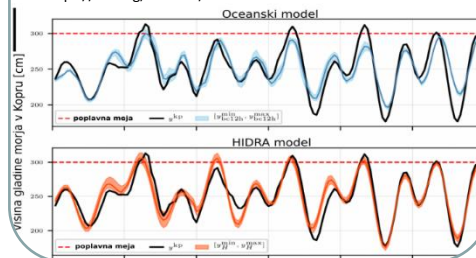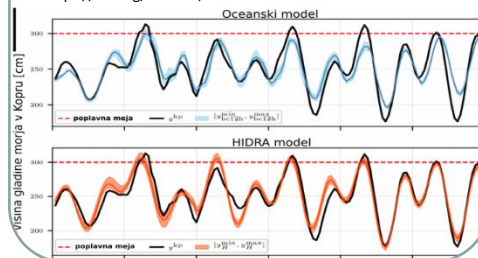
AI

Solving environment problems

Plastic pollution of see
https://www.sciencedirect.com/science/article/pii/S0269749121000683?via%3Dihub

Prediction of floods in coastal reagions
https://doi.org/10.3986/alternator.2021.35

Oceanski model

HIDRA model

Freedom vs. security – „predictive policing"
https://www.newscientist.com/article/2189986-a-uk-police-force-is-dropping-tricky-cases-on-advice-of-an-algorithm/

A UK police force uses an algorithm to choose which crimes to investigate. It has led to half as many assaults and public order offences being pursued.

However, because the technology bases its predictions on past investigations, any biases contained in those decisions may be reinforced by the algorithm.

Predictive maintenance

https://en.wikipedia.org/wiki/Predictive_maintenance

Predictive maintenance is a proactive maintenance technique that uses real-time asset data (collected through sensors), historical performance data, and advanced analytics to forecast when asset failure will occur.

THE REPUBLIC OF SLOVENIA
MINISTRY OF DIGITAL TRANSFORMATION

ITU Events https://aiforgood.itu.int/event/ai-powered-vehicles-for-humanitarian-help-deployment/

AI for Good Webinars

AI-powered vehicles for humanitarian help deployment

Thursday, 31 March 2022
16:00 - 18:00 Geneva (CEST)

In cooperation with
World Food Programme

aiforgood.itu.int

INNOVATION ACCELERATOR

Solving problems in medicine
https://www.nature.com/articles/s41598-019-51147-3.pdf

Our data demonstrate the feasibility of brain tumour diagnosis from routine blood tests using machine learning. the reported diagnostic accuracy is comparable and possibly complementary to that of imaging studies. the presented machine learning approach opens a completely new avenue in the diagnosis of these grave neurological diseases and demonstrates the utility of valuable information obtained from routine blood tests.

AI

Decision automation – „predictive criminal sentencing"
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
ProPublica found that the AI was not only often wrong, but also dangerously biased: it was more likely to rate Black defendants who did not reoffend as "high risk," and to rate white defendants who went on to reoffend as "low risk." The results showed that when an AI system is trained on historical data that reflects inequalities—as most data from the real world does—the system will project those inequalities into the future.

Freedom vs. security – „predictive policing"
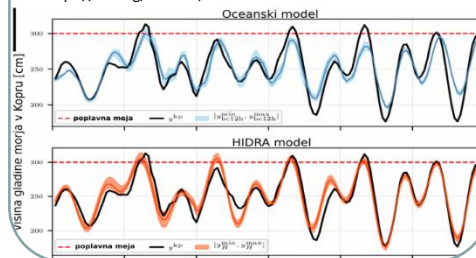https://www.newscientist.com/article/2189986-a-uk-police-force-is-dropping-tricky-cases-on-advice-of-an-algorithm/

A UK police force uses an algorithm to choose which crimes to investigate. It has led to half as many assaults and public order offences being pursued.

However, because the technology bases its predictions on past investigations, any biases contained in those decisions may be reinforced by the algorithm.

Solving environment problems

Plastic pollution of see
https://www.sciencedirect.com/science/article/pii/S0269749121000683?via%3Dihub

Prediction of floods in coastal reagions
https://doi.org/10.3986/alternator.2021.35

Oceanski model

HIDRA model

Predictive maintenance

https://en.wikipedia.org/wiki/Predictive_maintenance

Predictive maintenance is a proactive maintenance technique that uses real-time asset data (collected through sensors), historical performance data, and advanced analytics to forecast when asset failure will occur.

AI Omnipresence – Physical World

ITU Events https://aiforgood.itu.int/event/ai-powered-vehicles-for-humanitarian-help-deployment/

AI for Good Webinars

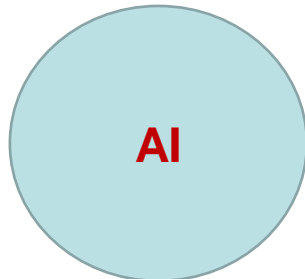**AI-powered vehicles for humanitarian help deployment**

Thursday, 31 March 2022
16:00 - 18:00 Geneva (CEST)

In cooperation with
World Food Programme

aiforgood.itu.int

INNOVATION ACCELERATOR

40 UN PARTNERS

„Deep fake" technologies
https://www.biometricupdate.com/201812/deep-fake-technology-outpacing-security-countermeasures

"A deep fake is the ability to manipulate sound images or video to make it appear that a certain person did something that they didn't do. These videos, in fact, are increasingly realistic. The quality of these fakes is rapidly increasing due to artificial intelligence [AI] machine learning algorithms paired with facial mapping software [that makes] it easy and cheap to insert someone's face into a video and produce a very realistic-looking video of someone saying or doing something they never said or did. This, by the way, technology is pretty widely available on the Internet, and people have used it already for all sorts of nefarious purposes at the individual level. I think you can only imagine what a nation-state could do with that technology, particularly to our politics."

Decision automation – „predictive criminal sentencing"
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
ProPublica found that the AI was not only often wrong, but also dangerously biased: it was more likely to rate Black defendants who did not reoffend as "high risk," and to rate white defendants who went on to reoffend as "low risk." The results showed that when an AI system is trained on historical data that reflects inequalities—as most data from the real world does—the system will project those inequalities into the future.
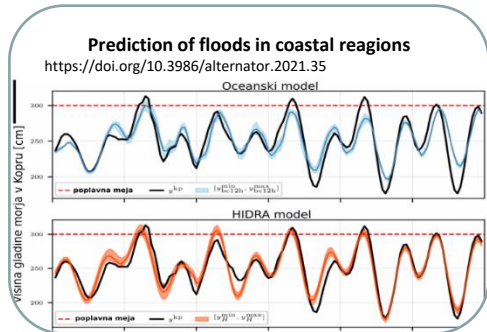
Solving problems in medicine
https://www.nature.com/articles/s41598-019-51147-3.pdf

Our data demonstrate the feasibility of brain tumour diagnosis from routine blood tests using machine learning. the reported diagnostic accuracy is comparable and possibly complementary to that of imaging studies. the presented machine learning approach opens a completely new avenue in the diagnosis of these grave neurological diseases and demonstrates the utility of valuable information obtained from routine blood tests.

AI

Freedom vs. security – „predictive policing"
https://www.newscientist.com/article/2189986-a-uk-police-force-is-dropping-tricky-cases-on-advice-of-an-algorithm/

A UK police force uses an algorithm to choose which crimes to investigate. It has led to half as many assaults and public order offences being pursued.

However, because the technology bases its predictions on past investigations, any biases contained in those decisions may be reinforced by the algorithm.

Solving environment problems

Plastic pollution of see
https://www.sciencedirect.com/science/article/pii/S0269749121000683?via%3Dihub

Prediction of floods in coastal reagions
https://doi.org/10.3986/alternator.2021.35

Oceanski model

HIDRA model

poplavna meja

Predictive maintenance

https://en.wikipedia.org/wiki/Predictive_maintenance

Predictive maintenance is a proactive maintenance technique that uses real-time asset data (collected through sensors), historical performance data, and advanced analytics to forecast when asset failure will occur.

AI Omnipresence – Physical World

ITU Events https://aiforgood.itu.int/event/ai-powered-vehicles-for-humanitarian-help-deployment/

AI for Good Webinars

AI-powered vehicles for humanitarian help deployment

Thursday, 31 March 2022
16:00 - 18:00 Geneva (CEST)

In cooperation with
World Food Programme

aiforgood.itu.int

INNOVATION ACCELERATOR

Solving problems in medicine
https://www.nature.com/articles/s41598-019-51147-3.pdf

Our data demonstrate the feasibility of brain tumour diagnosis from routine blood tests using machine learning. the reported diagnostic accuracy is comparable and possibly complementary to that of imaging studies. the presented machine learning approach opens a completely new avenue in the diagnosis of these grave neurological diseases and demonstrates the utility of valuable information obtained from routine blood tests.

Solving environment problems

Plastic pollution of see
https://www.sciencedirect.com/science/article/pii/S0269749121000683?via%3Dihub

Prediction of floods in coastal reagions
https://doi.org/10.3986/alternator.2021.35

AI

Ethical issues of AI
https://www.technologyreview.com/2021/01/29/1017065/ai-image-generation-is-racist-sexist/

Language-generation algorithms are known to embed racist and sexist ideas. ... Researchers have now demonstrated that the same can be true for image-generation algorithms. Feed one a photo of a man cropped right below his neck, and 43% of the time, it will autocomplete him wearing a suit. Feed the same one a cropped photo of a woman, even a famous woman like US Representative Alexandria Ocasio-Cortez, and 53% of the time, it will autocomplete her wearing a low-cut top or bikini. This has implications not just for image generation, but for all computer-vision applications, including video-based candidate assessment algorithms, facial recognition, and surveillance.

security-countermeasures

"A deep fake is the ability to manipulate sound images or video to make it appear that a certain person did something that they didn't do. These videos, in fact, are increasingly realistic. The quality of these fakes is rapidly increasing due to artificial intelligence [AI] machine learning algorithms paired with facial mapping software [that makes] it easy and cheap to insert someone's face into a video and produce a very realistic-looking video of someone saying or doing something they never said or did. This, by the way, technology is pretty widely available on the Internet, and people have used it already for all sorts of nefarious purposes at the individual level. I think you can only imagine what a nation-state could do with that technology, particularly to our politics."

Decision automation – „predictive criminal sentencing"
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

ProPublica found that the AI was not only often wrong, but also dangerously biased: it was more likely to rate Black defendants who did not reoffend as "high risk," and to rate white defendants who went on to reoffend as "low risk." The results showed that when an AI system is trained on historical data that reflects inequalities—as most data from the real world does—the system will project those inequalities into the future.

Freedom vs. security – „predictive policing"
https://www.newscientist.com/article/2189986-a-uk-police-force-is-dropping-tricky-cases-on-advice-of-an-algorithm/
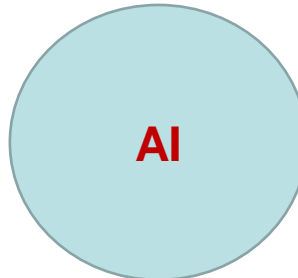
A UK police force uses an algorithm to choose which crimes to investigate. It has led to half as many assaults and public order offences being pursued.

However, because the technology bases its predictions on past investigations, any biases contained in those decisions may be reinforced by the algorithm.

Predictive maintenance

https://en.wikipedia.org/wiki/Predictive_maintenance

Predictive maintenance is a proactive maintenance technique that uses real-time asset data (collected through sensors), historical performance data, and advanced analytics to forecast when asset failure will occur.

AI Omnipresence – Physical World

ITU Events https://aiforgood.itu.int/event/ai-powered-vehicles-for-humanitarian-help-deployment/

AI for Good Webinars

AI-powered vehicles for humanitarian help deployment

Thursday, 31 March 2022
16:00 - 18:00 Geneva (CEST)

In cooperation with
World Food Programme

aiforgood.itu.int

INNOVATION ACCELERATOR

Ethical issues of AI
https://www.technologyreview.com/2021/01/29/1017065/ai-image-generation-is-racist-sexist/

... racist and sexist ideas. ... an be true for image-generation below his neck, and 43% of the time, ... one a cropped photo of a woman, ...dria Ocasio-Cortez, and 53% of the ... bikini. This has implications not just ...plications, including video-based candidate assessment algorithms, facial recognition, and surveillance.

Ethical issues of AI
https://www.abc.net.au/news/2021-07-17/anthony-bourdain-ai-voice-cloning-roadrunner/100301532

The revelation that a documentary filmmaker used voice-cloning software to make the late chef Anthony Bourdain say words he never spoke has drawn criticism amid ethical concerns about use of the powerful technology.

...security-countermeasures

"A deep fake is the ability to manipulate sound images or video to make it appear that a certain person did something that they didn't do. These videos, in fact, are increasingly realistic. The quality of these fakes is rapidly increasing due to artificial intelligence [AI] machine learning algorithms paired with facial mapping software [that makes] it easy and cheap to insert someone's face into a video and produce a very realistic-looking video of someone saying or doing something they never said or did. This, by the way, technology is pretty widely available on the Internet, and people have used it already for all sorts of nefarious purposes at the individual level. I think you can only imagine what a nation-state could do with that technology, particularly to our politics."

Solving problems in medicine
https://www.nature.com/articles/s41598-019-51147-3.pdf

Our data demonstrate the feasibility of brain tumour diagnosis from routine blood tests using machine learning. the reported diagnostic accuracy is comparable and possibly complementary to that of imaging studies. the presented machine learning approach opens a completely new avenue in the diagnosis of these grave neurological diseases and demonstrates the utility of valuable information obtained from routine blood tests.
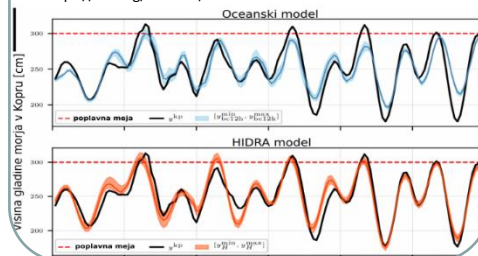
AI

Decision automation – „predictive criminal sentencing"
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

ProPublica found that the AI was not only often wrong, but also dangerously biased: it was more likely to rate Black defendants who did not reoffend as "high risk," and to rate white defendants who went on to reoffend as "low risk." The results showed that when an AI system is trained on historical data that reflects inequalities—as most data from the real world does—the system will project those inequalities into the future.

Freedom vs. security – „predictive policing"
https://www.newscientist.com/article/2189986-a-uk-police-force-is-dropping-tricky-cases-on-advice-of-an-algorithm/

A UK police force uses an algorithm to choose which crimes to investigate. It has led to half as many assaults and public order offences being pursued.

However, because the technology bases its predictions on past investigations, any biases contained in those decisions may be reinforced by the algorithm.

Solving environment problems

Plastic pollution of see
https://www.sciencedirect.com/science/article/pii/S0269749121000683?via%3Dihub

Prediction of floods in coastal reagions
https://doi.org/10.3986/alternator.2021.35

Predictive maintenance

https://en.wikipedia.org/wiki/Predictive_maintenance

Predictive maintenance is a proactive maintenance technique that uses real-time asset data (collected through sensors), historical performance data, and advanced analytics to forecast when asset failure will occur.

AI Omnipresence – Physical World

ITU Events https://aiforgood.itu.int/event/ai-powered-vehicles-for-humanitarian-help-deployment/

AI-powered humanita

Thursday, 31
16:00 - 18:00

In cooperation

aiforgood.itu.int

From prediction to shaping
https://www.theguardian.com/commentisfree/2019/jul/02/facebook-google-data-change-our-behaviour-democracy?mc_cid=9b16974e83&mc_eid=ae3284f75e

These experiences are translated into behavioural data. Some of this data may be applied to product or service improvements, and the rest is valued for its predictive power. These flows of predictive data are fed into computational products that predict human behavior. A leaked Facebook document in 2018 describes its machine-learning system that "ingests trillions of data points every day" and produces "more than 6m predictions per second". Finally, these prediction products are sold to business customers in markets that trade in human futures.

The idea is not only to know our behavior but also to shape it in ways that can turn predictions into guarantees. It is no longer enough to automate information flows about us; the goal now is to automate us.

Surveillance capitalists are rich and powerful, but they are not invulnerable. They fear law. They fear lawmakers. They fear citizens who insist on a different path. Both groups are bound together in the work of rescuing the digital future for democracy.

Solving prob
https://www.nature.com/article

Our data demonstrate the feasibility of brain tumour diagnosis from routine blood tests using machine learning. the reported diagnostic accuracy is comparable and possibly complementary to that of imaging studies. the presented machine learning approach opens a completely new avenue in the diagnosis of these grave neurological diseases and demonstrates the utility of valuable information obtained from routine blood tests.
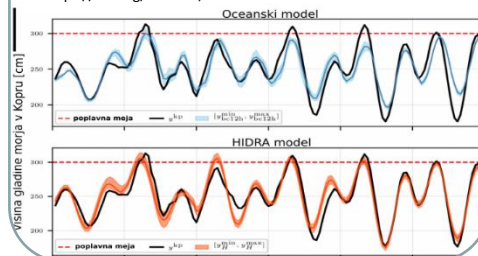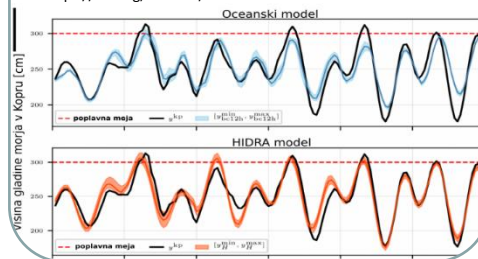
Ethical issues of AI
https://www.technologyreview.com/2021/01/29/1017065/ai-image-generation-is-racist-sexist/

Ethical issues of AI
https://www.abc.net.au/news/2021-07-17/anthony-bourdain-ai-voice-cloning-roadrunner/100301532

The revelation that a documentary filmmaker used voice-cloning software to make the late chef Anthony Bourdain say words he never spoke has drawn criticism amid ethical concerns about use of the powerful technology.

racist and sexist ideas. ...
an be true for image-generation
below his neck, and 43% of the time,
e one a cropped photo of a woman,
dria Ocasio-Cortez, and 53% of the
bikini. This has implications not just
lications, including video-based
candidate assessment algorithms, facial recognition, and surveillance.

security-countermeasures

"A deep fake is the ability to manipulate sound images or video to make it appear that a certain person did something that they didn't do. These videos, in fact, are increasingly realistic. The quality of these fakes is rapidly increasing due to artificial intelligence [AI] machine learning algorithms paired with facial mapping software [that makes] it easy and cheap to insert someone's face into a video and produce a very realistic-looking video of someone saying or doing something they never said or did. This, by the way, technology is pretty widely available on the Internet, and people have used it already for all sorts of nefarious purposes at the individual level. I think you can only imagine what a nation-state could do with that technology, particularly to our politics."

Decision automation – „predictive criminal sentencing"
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
ProPublica found that the AI was not only often wrong, but also dangerously biased: it was more likely to rate Black defendants who did not reoffend as "high risk," and to rate white defendants who went on to reoffend as "low risk." The results showed that when an AI system is trained on historical data that reflects inequalities—as most data from the real world does—the system will project those inequalities into the future.

Freedom vs. security – „predictive policing"
https://www.newscientist.com/article/2189986-a-uk-police-force-is-dropping-tricky-cases-on-advice-of-an-algorithm/

A UK police force uses an algorithm to choose which crimes to investigate. It has led to half as many assaults and public order offences being pursued.

However, because the technology bases its predictions on past investigations, any biases contained in those decisions may be reinforced by the algorithm.

Predictive maintenance

https://en.wikipedia.org/wiki/Predictive_maintenance

Predictive maintenance is a proactive maintenance technique that uses real-time asset data (collected through sensors), historical performance data, and advanced analytics to forecast when asset failure will occur.

AI

Solving environment problems

Plastic pollution of see
https://www.sciencedirect.com/science/article/pii/S0269749121000683?via%3Dihub

Prediction of floods in coastal reagions
https://doi.org/10.3986/alternator.2021.35

Oceanski model

HIDRA model

AI Omnipresence – Physical World

ITU Events https://aiforgood.itu.int/event/ai-powered-vehicles-for-humanitarian-help-deployment/

AI-powered humanita

Thursday, 31
16:00 - 18:00

In cooperation

aiforgood.itu.int

From prediction to shaping

https://www.theguardian.com/commentisfree/2019/jul/02/facebook-google-data-change-our-behaviour-democracy?mc_cid=9b16974e83&mc_eid=ae3284f75e

These experiences are translated into behavioural data. Some of this data may be applied to product or service improvements, and the rest is valued for its predictive power. These flows of predictive data are fed into computational products that predict human behavior. A leaked Facebook document in 2018 describes its machine-learning system that "ingests trillions of data points every day" and produces "more than 6m predictions per second". Finally, these prediction products are sold to business customers in markets that trade in human futures.

The idea is not only to know our behavior but also to shape it in ways that can turn predictions into guarantees. It is no longer enough to automate information flows about us; the goal now is to automate us.

Surveillance capitalists are rich and powerful, but they are not invulnerable. They fear law. They fear lawmakers. They fear citizens who insist on a different path. Both groups are bound together in the work of rescuing the digital future for democracy.

Ethical issues of AI
https://www.technologyreview.com/2021/01/29/1017065/ai-image-generation-is-racist-sexist/

racist and sexist ideas. ...
an be true for image-generation
below his neck, and 43% of the time,
one cropped photo of a woman,
dria Ocasio-Cortez, and 53% of the
bikini. This has implications not just
plications, including video-based
candidate assessment algorithms, facial recognition, and surveillance.

Ethical issues of AI
https://www.abc.net.au/news/2021-07-17/anthony-bourdain-ai-voice-cloning-roadrunner/100301532

The revelation that a documentary filmmaker used voice-cloning software to make the late chef Anthony Bourdain say words he never spoke has drawn criticism amid ethical concerns about use of the powerful technology.

security-countermeasures

"A deep fake is the ability to manipulate sound images or video to make it appear that a certain person did something that they didn't do. These videos, in fact, are increasingly realistic. The quality of these fakes is rapidly increasing due to artificial intelligence [AI] machine learning algorithms paired with facial mapping software [that makes] it easy and cheap to insert someone's face into a video and produce a very realistic-looking video of someone saying or doing something they never said or did. This, by the way, technology is pretty widely available on the Internet, and people have used it already for all sorts of nefarious purposes at the individual level. I think you can only imagine what a nation-state could do with that technology, particularly to our politics."

Solving prob
https://www.nature.com/articles/

Our data demonstrate the feasibility of brain tumour diagnosis from routine blood tests using machine learning.
the ren
comp
mach
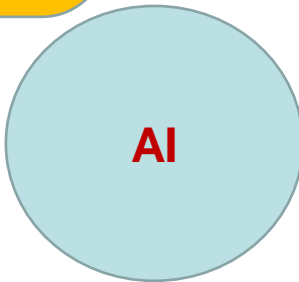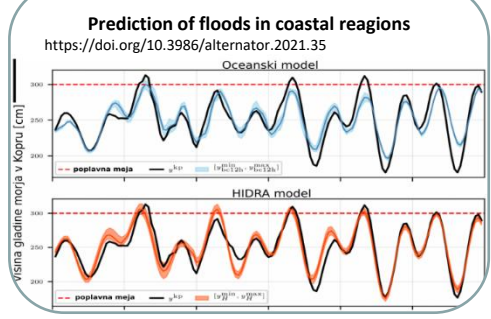in the
demo
from

Cambridge Analitica
https://knowledge.wharton.upenn.edu/article/fallout-cambridge-analytica/

"People are now beginning to realize that social media is not just either a fun plaything or a nuisance. It can have potentially real consequences in society."

The Cambridge Analytica scandal underscores how little consumers know about the potential uses of their data, according to Berman. He recalled a scene in the film *Minority Report* where Tom Cruise enters a mall and sees holograms of personally targeted ads. "Online advertising today has reached about the same level of sophistication, in terms of targeting, and also some level of prediction," he said. "It's not only that the advertiser can tell what you bought in the past, but also what you may be looking to buy."

https://www.sciencedirect.com/science/article/pii/S0269749121000683?via%3Dihub

AI

Decision automation – „predictive criminal sentencing"
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
ProPublica found that the AI was not only often wrong, but also dangerously biased: it was more likely to rate Black defendants who did not reoffend as "high risk," and to rate white defendants who went on to reoffend as "low risk." The results showed that when an AI system is trained on historical data that reflects inequalities—as most data from the real world does—the system will project those inequalities into the future.

Freedom vs. security – „predictive policing"
https://www.newscientist.com/article/2189986-a-uk-police-force-is-dropping-tricky-cases-on-advice-of-an-algorithm/

A UK police force uses an algorithm to choose which crimes to investigate. It has led to half as many assaults and public order offences being pursued.

However, because the technology bases its predictions on past investigations, any biases contained in those decisions may be reinforced by the algorithm.

Predictive maintenance

https://en.wikipedia.org/wiki/Predictive_maintenance

Predictive maintenance is a proactive maintenance technique that uses real-time asset data (collected through sensors), historical performance data, and advanced analytics to forecast when asset failure will occur.

Prediction of floods in coastal reagions
https://doi.org/10.3986/alternator.2021.35

Oceanski model

Višina gladine morja v Kopru [cm]

poplavna meja

HIDRA model

poplavna meja

AI Omnipresence – Physical World

ITU **Events** https://aiforgood.itu.int/event/ai-powered-vehicles-for-humanitarian-help-deployment/

AI-powered
humanita

Thursday, 31
16:00 - 18:00

In cooperation

aiforgood.itu.

From prediction to shaping

https://www.theguardian.com/commentisfree/2019/jul/02/facebook-google-data-change-our-behaviour-democracy?mc_cid=9b16974e83&mc_eid=ae3284f75e

These experiences are translated into behavioural data. Some of this data may be applied to product or service improvements, and the rest is valued for its predictive power. These flows of predictive data are fed into computational products that predict human behavior. A leaked Facebook document in 2018 describes its machine-learning system that "ingests trillions of data points every day" and produces "more than 6m predictions per second". Finally, these prediction products are sold to business customers in markets that trade in human futures.

The idea is not only to know our behavior but also to shape it in ways that can turn predictions into guarantees. It is no longer enough to automate information flows about us; the goal now is to automate us.

Surveillance capitalists are rich and powerful, but they are not invulnerable. They fear law. They fear lawmakers. They fear citizens who insist on a different path. Both groups are bound together in the work of rescuing the digital future for democracy.

Solving prob
https://www.nature.com/article

Our data demonstrate the feasibility of brain tumour diagnosis from routine blood tests using machine learning.
the rer
comp
mach
in the
demo
from

Cambridge Analitica
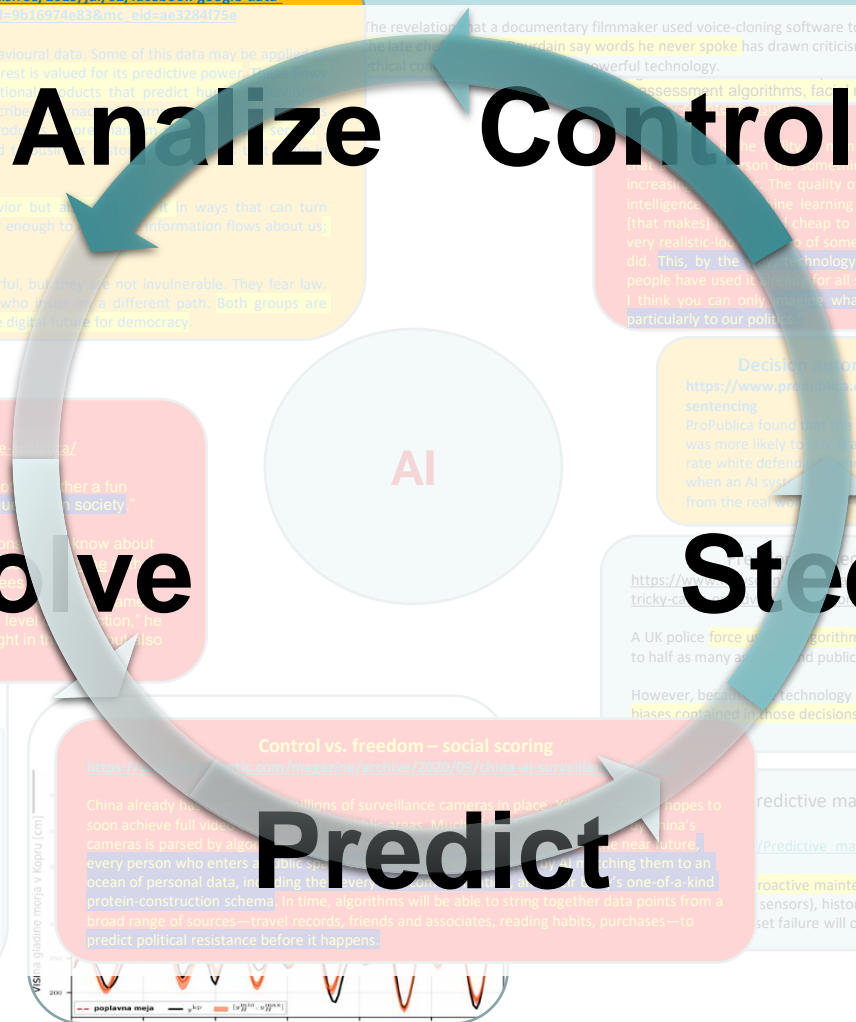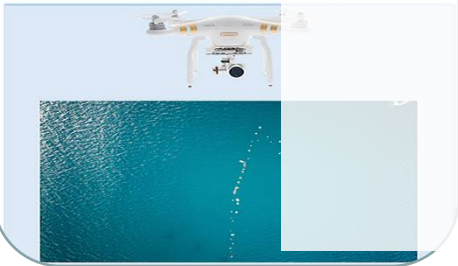https://knowledge.wharton.upenn.edu/article/fallout-cambridge-analytica/

"People are now beginning to realize that social media is not just either a fun plaything or a nuisance. It can have potentially real consequences in society."

The Cambridge Analytica scandal underscores how little consumers know about the potential uses of their data, according to Berman. He recalled a scene in the film *Minority Report* where Tom Cruise enters a mall and sees holograms of personally targeted ads. "Online advertising today has reached about the same level of sophistication, in terms of targeting, and also some level of prediction," he said. "It's not only that the advertiser can tell what you bought in the past, but also what you may be looking to buy."

https://www.sciencedirect.com/science/article/pii/S02697
49121000683?via%3Dihub

Ethical issues of AI
https://www.technologyreview.com/2021/01/29/1017065/ai-image-generation-is-racist-sexist/

Ethical issues of AI
https://www.abc.net.au/news/2021-07-17/anthony-bourdain-ai-voice-cloning-roadrunner/100301532

The revelation that a documentary filmmaker used voice-cloning software to make the late chef Anthony Bourdain say words he never spoke has drawn criticism amid ethical concerns about use of the powerful technology.

racist and sexist ideas. ...
an be true for image-generation
elow his neck, and 43% of the time,
e one a cropped photo of a woman,
dria Ocasio-Cortez, and 53% of the
bikini. This has implications not just
ications, including video-based
candidate assessment algorithms, facial recognition, and surveillance.

security-countermeasures

"A deep fake is the ability to manipulate sound images or video to make it appear that a certain person did something that they didn't do. These videos, in fact, are increasingly realistic. The quality of these fakes is rapidly increasing due to artificial intelligence [AI] machine learning algorithms paired with facial mapping software [that makes] it easy and cheap to insert someone's face into a video and produce a very realistic-looking video of someone saying or doing something they never said or did. This, by the way, technology is pretty widely available on the Internet, and people have used it already for all sorts of nefarious purposes at the individual level. I think you can only imagine what a nation-state could do with that technology, particularly to our politics."

Decision automation – „predictive criminal sentencing"
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

ProPublica found that the AI was not only often wrong, but also dangerously biased: it was more likely to rate Black defendants who did not reoffend as "high risk," and to rate white defendants who went on to reoffend as "low risk." The results showed that when an AI system is trained on historical data that reflects inequalities—as most data from the real world does—the system will project those inequalities into the future.

Freedom vs. security – „predictive policing"
https://www.newscientist.com/article/2189986-a-uk-police-force-is-dropping-tricky-cases-on-advice-of-an-algorithm/

A UK police force uses an algorithm to choose which crimes to investigate. It has led to half as many assaults and public order offences being pursued.

However, because the technology bases its predictions on past investigations, any biases contained in those decisions may be reinforced by the algorithm.

Control vs. freedom – social scoring
https://www.theatlantic.com/magazine/archive/2020/09/china-ai-surveillance/614197/

China already has hundreds of millions of surveillance cameras in place. Xi's government hopes to soon achieve full video coverage of key public areas. Much of the footage collected by China's cameras is parsed by algorithms for security threats of one kind or another. In the near future, every person who enters a public space could be identified, instantly, by AI matching them to an ocean of personal data, including their every text communication, and their body's one-of-a-kind protein-construction schema. In time, algorithms will be able to string together data points from a broad range of sources—travel records, friends and associates, reading habits, purchases—to predict political resistance before it happens.

predictive maintenance

/Predictive_maintenance

roactive maintenance technique that uses real-time
sensors), historical performance data, and advanced
et failure will occur.

AI

Višina gladine morja v Kopru [cm]
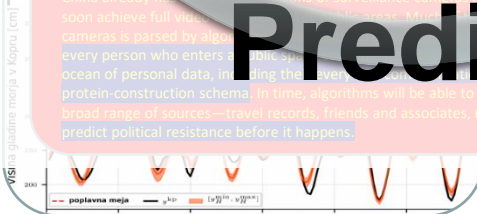
poplavna meja

Analize **Control** **Steer** **Predict** **Solve**

AI

ITU Events https://aiforgood.itu.int/event/ai-powered-vehicles-for-humanitarian-help-deployment/

AI-powered humanita...

Thursday, 31
16:00 - 18:00

In cooperation

aiforgood.itu.in

From prediction to shaping
https://www.theguardian.com/commentisfree/2019/jul/02/facebook-google-data-change-our-behaviour-democracy?mc_cid=9b16974e83&mc_eid=ae3284f75e

These experiences are translated into behavioural data. Some of this data may be applied to product or service improvements, and the rest is valued for its predictive power. These flows of predictive data are fed into computational products that predict human behavior. A leaked Facebook document in 2018 described its machine-learning system that "ingests trillions of data points every day" and produces "more than 6 million predictions per second." Finally, these prediction products are sold to business customers... that trade in human futures.

The idea is not only to know our behavior but also to shape it in ways that can turn predictions into guarantees. It is no longer enough to automate information flows about us; the goal now is to automate us.

Surveillance capitalists are rich and powerful, but they are not invulnerable. They fear law. They fear lawmakers. They fear citizens who insist on a different path. Both groups are bound together in the work of rescuing the digital future for democracy.

Solving prob...
https://www.nature.com/article...

Our data demonstrate the feasibility of brain tumour diagnosis from routine blood tests using machine learning...

https://www.sciencedirect.com/science/article/pii/S0269749121000683?via%3Dihub

Cambridge Analitica
https://knowledge.wharton.upenn.edu/article/fallout-cambridge-analytica/

"People are now beginning to realize that social media is no longer either a fun plaything or a nuisance. It can have potentially real consequences on society."

The Cambridge Analytica scandal underscores... people concerned... know about the potential uses of their data, according to Be... film *Minority Report* where Tom Cruise enters a mall... personally targeted ads. "Online advertising to... level of sophistication, in terms of targeting, and also some level... ion," he said. "It's not only that the advertiser can tell what you bought in the past but also what you may be looking to buy."

Ethical issues of AI
https://www.technologyreview.com/2021/01/29/1017065/ai-image-generation-is-racist-sexist/

... racist and sexist ideas. ...
... an be true for image-generation
... below his neck, and 43% of the time,
... one cropped photo of a woman,"
... dria Ocasio-Cortez, and 53% of the
... bikini. This has implications not just
... plications, including video-based
... assessment algorithms, facial recognition, and surveillance.

Ethical issues of AI
https://www.abc.net.au/news/2021-07-17/anthony-bourdain-ai-voice-cloning-roadrunner/100301532

The revelation that a documentary filmmaker used voice-cloning software to make the late chef... Bourdain say words he never spoke has drawn criticism amid ethical con...
... powerful technology.

... nipulate sound images or video to make it appear... person doing something that they didn't do. These videos, in fact, are increasing... The quality of these fakes is rapidly increasing due to artificial intelligence... machine learning algorithms paired with facial mapping software [that makes]... cheap to insert someone's face into a video and produce a very realistic-looking video of someone saying or doing something they never said or did. This, by the... technology is pretty widely available on the Internet, and people have used it... for all sorts of nefarious purposes at the individual level. I think you can only imagine what a nation-state could do with that technology, particularly to our politics.

Decision automation – „predictive criminal sentencing"
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

ProPublica found that the AI was not only often wrong, but also dangerously biased: it was more likely to rate Black defendants who did not reoffend as "high risk," and to rate white defendants who did go on to reoffend as "low risk." The results showed that when an AI system is trained on historical data that reflects inequalities—as most data from the real world does—the system will project those inequalities into the future.

... predictive policing"
https://www... 86-a-uk-police-force-is-dropping-tricky-c...

A UK police force used an algorithm to choose which crimes to investigate. It has led to half as many... and public order offences being pursued.

However, because the technology bases its predictions on past investigations, any biases contained in those decisions may be reinforced by the algorithm.

Control vs. freedom – social scoring
...https://...com/magazine/archive/2020/09/china-ai-surveillance/...

China already has... millions of surveillance cameras in place... hopes to soon achieve full vide... cameras is parsed by algo... near future... every person who enters a public space... ching them to an ocean of personal data, in... ... one-of-a-kind protein-construction schema. In time, algorithms will be able to string together data points from a broad range of sources—travel records, friends and associates, reading habits, purchases—to predict political resistance before it happens.

... predictive maintenance
...https://...Predictive_maintenance

...roactive maintenance technique that uses real-time... sensors), historical performance data, and advanced... ...et failure will occur.

GAI

Large Language Models

Table 1. Overview of existing LLMs

Model	Year of Release	Developer (Country)	Transformer Type	Number of Parameters (Billions)
Wu Dao 2.0	2021	BAAI (China)	Multimodal	1750
GLaM	2021	GOOGLE (US)	Seq2seq	1200
MT-NLG	2021	Microsoft, Nvidia (US)	Seq2seq	530
Gopher	2021	DeepMind (US)	Autoregressive	280
HyperCLOVA	2021	Naver Corporation (Korea)	Autoregressive	204
BLOOM	2022	Hugging Face (U.S.)	Autoregressive	176
OPT-175B	2022	Meta (U.S.)	Autoregressive	176
GPT-3	2020	OpenAI (U.S.)	Autoregressive	175
YaLM 100B	2022	Yandex (Russia)	Autoregressive	100
AlexaTM 20B	2022	Amazon (U.S.)	Seq2seq	20

© **The AI Natural Language Processing Landscape, Working Party on Artificial Intelligence Governance, OECD, 2022**

AI Omnipresence – Digital World

Table 1. Overview of existing LLMs

Model	Year of Release	Developer (Country)	Transformer Type	Number of Parameters (Billions)
Wu Dao 2.0	2021	BAAI (China)	Multimodal	1750
GLaM	2021	GOOGLE (US)	Seq2seq	1200
MT-NLG	2021	Microsoft, Nvidia (US)	Seq2seq	530
Gopher	2021	DeepMind (US)	Autoregressive	280
HyperCLOVA	2021	Naver Corporation (Korea)	Autoregressive	204
BLOOM	2022	Hugging Face (U.S.)	Autoregressive	176
OPT-175B	2022	Meta (U.S.)	Autoregressive	176
GPT-3	2020	OpenAI (U.S.)	Autoregressive	175
YaLM 100B	2022	Yandex (Russia)	Autoregressive	100
AlexaTM 20B	2022	Amazon (U.S.)	Seq2seq	20

© **The AI Natural Language Processing Landscape, Working Party on Artificial Intelligence Governance, OECD, 2022**

THE REPUBLIC OF SLOVENIA
MINISTRY OF DIGITAL TRANSFORMATION

Developer (Country)	Transformer Type	Number of Parameters (Billions)
BAAI (China)	Multimodal	1750
GOOGLE (US)	Seq2seq	1200
Microsoft, Nvidia (US)	Seq2seq	530
DeepMind (US)	Autoregressive	280
Naver Corporation (Korea)	Autoregressive	204
Hugging Face (U.S.)	Autoregressive	176
Meta (U.S.)	Autoregressive	176
OpenAI (U.S.)	Autoregressive	175
Yandex (Russia)	Autoregressive	100
Amazon (U.S.)	Seq2seq	20

Working Party on Artificial Intelligence Governance, OECD, 2022

AI Omnipresence – Digital World

https://marketrealist.com/2015/12/adoption-rates-dizzying-heights/

https://www.gartner.com/en/articles/the-4-trends-that-prevail-on-the-gartner-hype-cycle-for-ai-2021

https://ai-from-ambition-to-action.com/

Asistent 86   Learning of decision trees

https://ai-from-ambition-to-action.com/

THE REPUBLIC OF SLOVENIA
MINISTRY OF DIGITAL TRANSFORMATION

Stephen Hawking warned AI could serve as the "worst event in the history of our civilization" unless humanity is prepared for its possible risks.

Elon Musk has been outspoken about the potential dangers of AI and the need for government regulation. During the National Governors Association meeting in July, Musk urged lawmakers to consider rules for how AI is created.
"AI is a rare case where I think we need to be proactive in regulation than reactive," said Musk.
"Nobody likes being regulated, but everything (cars, planes, food, drugs, etc) that's a danger to the public is regulated," said Musk on Twitter. "AI should be too.„

Tim Berners-Lee: 'The system is failing'
The spread of misinformation and propaganda online has exploded partly because of the way the advertising systems of large digital platforms such as Google or Facebook have been designed to hold people's attention.
"People are being distorted by very finely trained AIs that figure out how to distract them," said Berners-Lee.

Future of Life open letter (Yoshua Bengio, Stuart Russel): Ssystems with human-competitive intelligence can pose profound risks to society and humanity, as shown by extensive research[1] and acknowledged by top AI labs.[2] As stated in the widely-endorsed Asilomar AI Principles, *Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources*. ...Therefore, **we call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4**.

Geoffrey Hinton: Geoffrey Hinton, 75, announced his resignation from Google in a statement to the New York Times, saying he now regretted his work.
He told the BBC some of the dangers of AI chatbots were "quite scary".
"Right now, they're not more intelligent than us, as far as I can tell. But I think they soon may be."

NpUI

AI system definition

EU HLEG ethical guidelines

Framework for Trustworthy AI

INTRODUCTION

Trustworthy AI

| **Lawful AI** | **Ethical AI** | **Robust AI** |

(not dealt with in this document)

CHAPTER I

Foundations of Trustworthy AI

Adhere to ethical principles based on fundamental rights

→ **4 Ethical Principles**

Acknowledge and address tensions between them

→
- Respect for human autonomy
- Prevention of harm
- Fairness
- Explicability

CHAPTER II

Realisation of Trustworthy AI

Implement the key requirements

→ **7 Key Requirements**

Evaluate and address these continuously throughout the AI system's life cycle via

Technical Methods **Non-Technical Methods**

→
- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental wellbeing
- Accountability

NpUI – AI system (OECD)

AI System

Sensors: (§1)
- Machine (§2)
- Human (§3)

Processed by Machine

Structured Data

Processed by Human

AI Operational Logic

Model Construction Algorithm (§4) (e.g., machine learning)

Model (§8) (e.g., rules or analytical function)

Model Interpretation Algorithm (§10) (e.g., classification or logic reasoning)

Objective (§11)

Performance Measure (§12)

Objective (§5)

Performance Measure (§6)

Historical data / memory (§7)

Human Interpretable or Uninterpretable Representation (§9)

Recommendations

Actuators: (§13)
- Machine (§14)
- Human (§15)

Environment

Perceiving (Percepts / Raw Data)

Environment State
- An environment has its state, not necessarily fully observable
- An environment can change its state with or without explicit actions by the AI System

Environment State Observability
- Environments are real (physical, social, mental) or artificial (e.g., games like chess)
- Real environments are typically too complex to be observable in its entirety (like physical or social)
- Artificial environments can be fully observable (like chess)
- Environments are observable through 'percepts' in a form of raw data

Acting (Physical or Informational Influence)

NpUI – AI system (OECD)

AI System

Environment

Sensors: (§1)
- Machine (§2)
- Human (§3)

Structured Data

Processed by Machine

Processed by Human

AI Operational Logic

Model Construction Algorithm (§4) (e.g., machine learning)

Model (§8) (e.g., rules or analytical function)

Model Interpretation Algorithm (§10) (e.g., classification or logic reasoning)

Objective (§11)

Performance Measure (§12)

Objective (§5)

Performance Measure (§6)

Historical data / memory (§7)

Human Interpretable or Uninterpretable Representation (§9)

Recommendations

Actuators: (§13)
- Machine (§14)
- Human (§15)

Perceiving (Percepts / Raw Data)

Data bias
- non-representative data (representation error)
- low quality data (measurement error)
- historic data (outdated data)

Environment State
- An environment has its state, not necessarily fully observable
- An environment can change its ... without explicit ... System
- Environments are real (physical, social, mental) or artificial (e.g., games like chess)
- Real environments are typically too complex to be observable in its entirety (like physical or social)
- Artificial environments can be fully observable (like chess)
- Environments are observable through 'percepts' in a form of raw data

Acting (Physical or Informational Influence)

NpUI – AI system (OECD)

AI System

Environment

Sensors: (§1)
- Machine (§2)
- Human (§3)

Perceiving (Percepts / Raw Data)

Environment State
- An environment has its state, not necessarily fully observable
- An environment can change its ... out explicit ... System

Processed by Machine

Structured Data

Processed by Human

Data bias
- non-representative data (representation error)
- low quality data (measurement error)
- historic data (outdated data)

AI Operational Logic

Model Construction Algorithm (§4) (e.g., machine learning)

Model (§8) (e.g., rules or analytical function)

Model Interpretation Algorithm (§10) (e.g., classification or logic reasoning)

Objective (§11)

Performance Measure (§12)

- Environments are real (physical, social, mental) or artificial (e.g., games like chess)
- Real environments are typically too complex to be observable in its entirety (like physical or social)
- Artificial environments can be fully observable (like chess)
- Environments are observable through 'percepts' in a form of raw data

Objective (§5)

Performance Measure (§6)

Historical data / memory (§7)

Human Interpretable or Uninterpretable Representation (§9)

Recommendations

Actuators: (§13)
- Machine (§14)
- Human (§15)

Acting (Physical or Informational Influence)

Bias in algorithms

NpUI – AI system (OECD)

AI System

Environment

Sensors: (§1)
- Machine (§2)
- Human (§3)

Processed by Machine

Processed by Human

Structured Data

Perceiving (Percepts / Raw Data)

Data bias
- non-representative data (representation error)
- low quality data (measurement error)
- historic data (outdated data)

AI Operational Logic

Model Construction Algorithm (§4) (e.g., machine learning)

Model (§8) (e.g., rules or analytical function)

Model Interpretation Algorithm (§10) (e.g., classification or logic reasoning)

Objective (§11)

Performance Measure (§12)

Objective (§5)

Performance Measure (§6)

Historical data / memory (§7)

Human Interpretable or Uninterpretable Representation (§9)

Recommendations

Actuators: (§13)
- Machine (§14)
- Human (§15)

Acting (Physical or Informational Influence)

Environment State
- An environment has its state, not necessarily fully observable
- An environment can change its ... without explicit ... System
- Environments are real (physical, social, mental) or artificial (e.g., games like chess)
- Real environments are typically too complex to be observable in its entirety (like physical or social)
- Artificial environments can be fully observable (like chess)
- Environments are observable through 'percepts' in a form of raw data

Bias in algorithms

Interpretation bias

Some examples of biased results
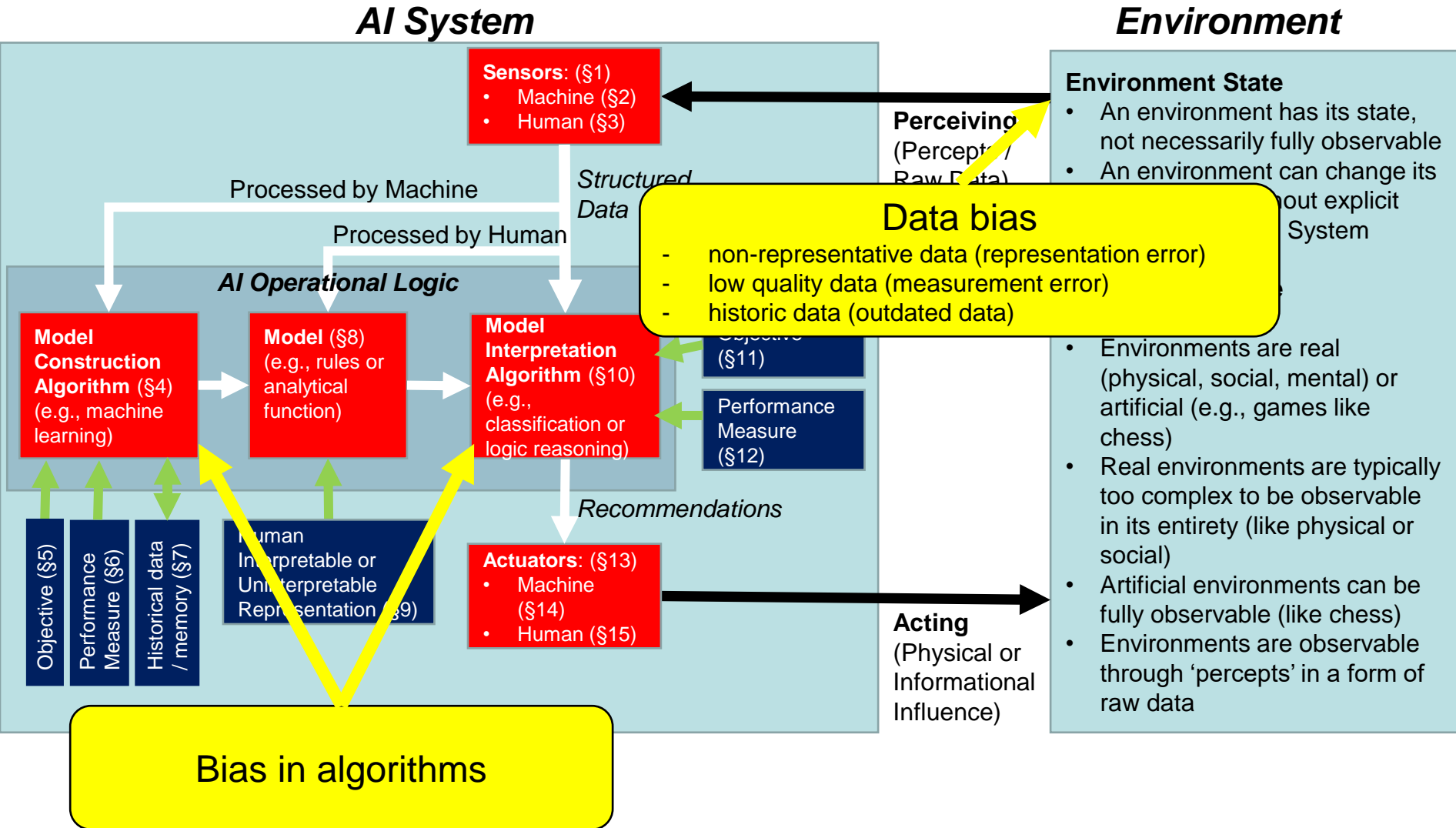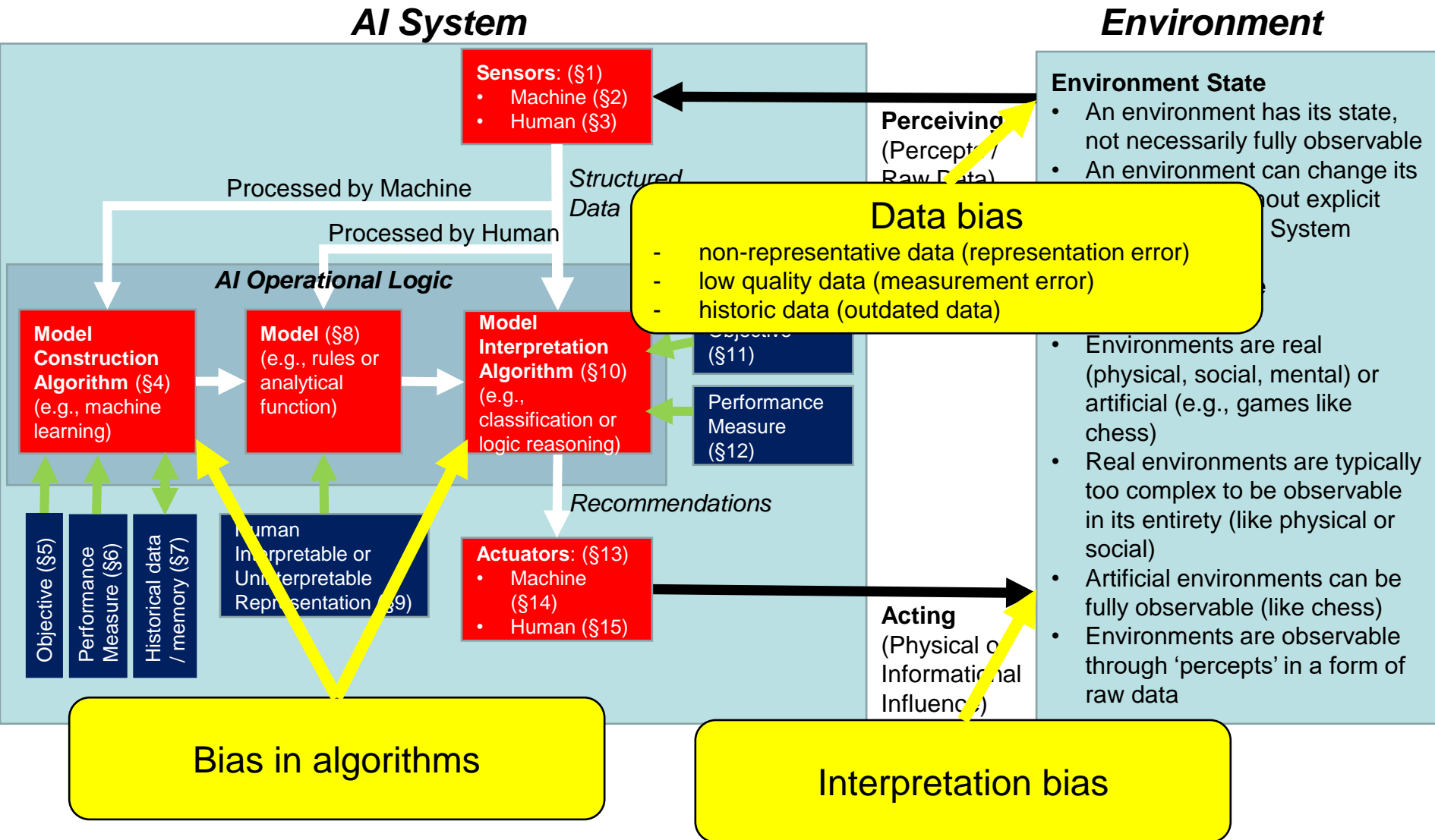
- Gender bias in language recognition/generation
- Racial/gender bias in biometric facial/voice/etc. recognition
- Racial bias in justice systems
- Bias in predictive analitics (employement, financial fraud, loan administration, etc.)
- Bias in social services and welfare

FRA:

- Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights (https://fra.europa.eu/en/publication/2019/data-quality-and-artificial-intelligence-mitigating-bias-and-error-protect)
- #BigData: Discrimination in data-supported decision making (https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making)

NpUI

concept, context, strategic objectives

THE REPUBLIC OF SLOVENIA
MINISTRY OF DIGITAL TRANSFORMATION

NpUI – what and how

- **reference point for understanding** the issues **and planning** AI activities for all stakeholders involved at the national level;
- **coherent proposal for systemic cross-sectoral support**, regulation and implementation of all activities related to AI in Slovenia;
- **governance structure** that will be able to **dynamically plan and coordinate the implementation** of all envisaged activities, which are by their very nature **diverse and cross-sectorial**;
- **concrete** objectives, measures, implementation instruments and finances;

NpUI – concept

- **„glocal" ecosystem approach** – dynamic, quick, adaptive;
- **innovation diffusion lifecycle approach** tailored to Slovenian needs.

NpUI – Vision

Upgrade more than 40 years of research achievements i**n the field of AI in Slovenia and become internationally recognised for the competences of knowledge transfer of high-quality, ethical and safe AI technologies in human-friendly and trustworthy services and products while ensuring national cultural identity.**

By fully supporting Slovenian research & innovation stakeholders in the development of AI based technologies and solutions, by introducing and implementing reference solutions based on AI in cooperation with all social groups in Slovenia, and by supporting the implementation of Slovenian stakeholders in the field of AI, we want to. establish Slovenia as a credible partner in the implementation and regulation of AI in society, based on human-centric approach.

NpUI – Strategic Goals/Policy Areas

1. Setting up an dynamic ecosystem for research, innovation and deployment of AI
2. Education and strengthening of human resources
3. Support for AI research and innovation
4. Introduction of reference AI implementations into economy, public sector, public administration and society
5. Establishment of infrastructure for research, development and use of AI
6. Enhancing security by using AI
7. Increasing public confidence in the AI
8. Provide an appropriate legal and ethical framework
9. Strengthening international cooperation
10. Establishment of a national Observatory for AI

NpUI – Areas of Action

Actions: in-direct support	Actions: direct support

Preparation and response to changes

- E-skills and e-competences for AI
- Ethics for the AI
- Regulation of the AI
- Public confidence in the AI

AI RDI

Deployment of AI

Priority areas:
- Language Technologies
- eHealth
- Industry 4.0
- Public administration
- Spatial planning
- Sustainable food & environment

AI infrastructure

International cooperation

Standardisation

1. Setting up an dynamic ecosystem for research, innovation and deployment of AI
2. Education and strengthening of human resources
3. Support for AI research and innovation
4. Introduction of reference AI implementations into economy, public sector, public administration and society
5. Establishment of infrastructure for research, development and use of AI
6. Enhancing security by using AI
7. Increasing public confidence in the AI
8. Provide an appropriate legal and ethical framework
9. Strengthening international cooperation
10. Establishment of a national Observatory for AI

1. Setting up an dynamic ecosystem for research, innovation and deployment of AI
2. Education and strengthening of human resources
3.
4.
5.
6.
7.
8.
9.
10

2.4	An analysis of the needs and possibilities of developing **interdisciplinary secondary and higher education study programmes linking AI and data science on the one hand with humanities and law** on the other;
2.6	The preparation of a **platform and educational content for distance education** at all levels of formal education and lifelong learning with the aim of enabling the acquisition of advanced professional digital skills, in particular in the field of AI and data science and the social, **ethical and legal aspects** of the AI;
2.8	**Analysis of labour market transformation**, working relationships, working conditions and labour organisation, employment profiles and jobs in various sectors in Slovenia that have the potential to be replaced or modified as a result of the introduction of the AI, including from the point of view of **gender equality**, over a period of 10 years;

1. Setting up an dynamic ecosystem for research, innovation and deployment of AI
2. Education and strengthening of human resources
3. Support for AI research and innovation
4. Introduction of reference AI implementations into economy, public sector, public administration and society
5. Establishment of infrastructure for research, development and use of AI
6. Enhancing security by using AI
7. Increasing public confidence in the AI
8. Provide an appropriate legal and ethical framework
9. Strengthening international cooperation
10. Establishment of a national Observatory for AI

1. Setting up an dynamic ecosystem for research, innovation and deployment of AI
2. Education and strengthening of human resources
3. Support for AI research and innovation
4. Introduction of reference AI implementations into economy, public sector, public administration and society
5. Establishment of infrastructure for research, development and use of AI
6.
7.

| **5.4** | Analyzing mechanisms and **defining a legal ethical framework for governance** (collection, storage, access, use, modification, etc.) **of non-personal data** within and between the industry, the public sector and the research sphere (including in terms of the right to privacy) in line with and co-operate with relevant activities at EU level; |

8.
9.
10. Establishment of a national Observatory for AI

1. Setting up an dynamic ecosystem for research, innovation and deployment of AI
2. Education and strengthening of human resources
3. Support for AI research and innovation
4. Introduction of reference AI implementations into economy, public sector, public administration and society
5. Establishment of infrastructure for research, development and use of AI
6. Enhancing security by using AI
7. Increasing public confidence in the AI
8. Provide an appropriate legal and ethical framework
9. Strengthening international cooperation
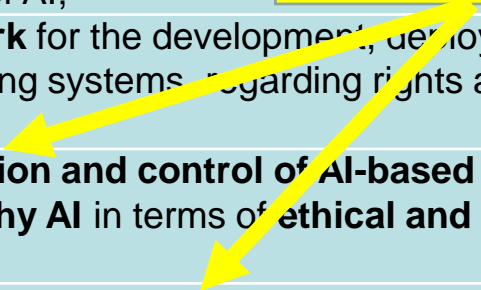10. Establishment of a national Observatory for AI

1. Setting up an dynamic ecosystem for research, innovation and deployment of AI
2. Education and strengthening of human resources
3. Support for AI research and innovation
4. Introduction of reference AI implementations into economy, public sector, public administration and society
5. Establishment of infrastructure for research, development and use of AI
6. Enhancing security by using AI
7. Increasing public confidence in the AI
8.
9.
10.

7.5	**Defining a public confidence framework for AI solutions** based on the link between the technological characteristics of AI technologies, methodologies for developing and applying AI solutions, standardisation and use of FLOSS software on the one hand, **and ethical and legal principles on the other**.
7.6	Support to **NGO networking and coordination for participation in activities of research, development, deployment and use of AI** at national and EU level.
7.7	Promoting **NGO projects** to ensure public confidence in AI.

NpUI – SO8

1. Setting up an dynamic ecosystem for research, innovation and deployment of AI
2. Education and strengthening of human resources
3. Support for AI research and innovation
4. Introduction of reference AI implementations into economy, public sector, public administration and society
5. Establishment of infrastructure for research, development and use of AI
6. Enhancing security by using AI
7. Increasing public confidence in the AI
8. Provide an appropriate legal and ethical framework
9. Strengthening international cooperation
10. Establishment of a national Observatory for AI

8.1	**Inclusion of humanities, social sciences** and especially legal and security experts and representatives of NGOs in activities in the field of AI, in the support pillar of the Slovenian Digital Coalition;
8.2	Active monitoring and involvement in the **discussions on the development of AI in international organisations** (e.g. EU, UNESCO, Council of Europe) **dealing with ethics and law in the AI** field and working towards the adoption of an international umbrella convention that would constitute legal standards for the protection of human rights in the development, deployment and application of AI;
8.3	**An analysis of the legal ethical framework** for the development, deployment and use of AI-based systems, especially in decision-making systems regarding rights and duties of natural and legal persons;
8.4	Establishment of a framework for **certification and control of AI-based solutions** in line with the EU framework for ensuring a **trustworthy AI** in terms of **ethical and robust requirements**;
8.5	Establishment of a **national supervisory mechanism** to monitor and verify the compliance of AI-based solutions with EU legislation providing a trusted AI;
8.6	The organisation of an **annual conference on the legal ethical framework** for the development, deployment and application of AI in various areas (e.g. the fight against crime, autonomous vehicles, health, taxes);
8.7	**Cooperation with centres of knowledge** (e.g. UNESCO International Centre for Research of the AI) **on legal, social and ethical issues of AI**.The participation of scientists in the field of humanities and social sciences and representatives of NGOs shall be included.

|	
8.1	**Inclusion of humanities, social sciences** and especially legal and security experts and representatives of NGOs in activities in the field of AI, in the support pillar of the Slovenian Digital Coalition;
8.2	Active monitoring and involvement in the **discussions on th... international organisations** (e.g. EU, UNESCO, Council of ...thics **and law in the AI** field and working towards the adoption of ... convention that would constitute legal standards for the prote... the development, deployment and application of AI;
8.3	**An analysis of the legal ethical framework** for the development, deployment and use of AI-based systems, especially in decision-making systems regarding rights and duties of natural and legal persons;
8.4	Establishment of a framework for **certification and control of AI-based solutions** in line with the EU framework for ensuring a **trustworthy AI** in terms of **ethical and robust requirements**;
8.5	Establishment of a **national supervisory mechanism** to monitor and verify the compliance of AI-based solutions with EU legislation providing a trusted AI;
8.6	The organisation of an **annual conference on the legal ethical framework** for the development, deployment and application of AI in various areas (e.g. the fight against crime, autonomous vehicles, health, taxes);
8.7	**Cooperation with centres of knowledge** (e.g. UNESCO International Centre for Research of the AI) **on legal, social and ethical issues of AI**.The participation of scientists in the field of humanities and social sciences and representatives of NGOs shall be included.

AI Act

AI Act

Requirements:

- Risk management systems
- Data and data governance
- Technical documentation
- Record-keeping
- Transparency and provision of information to users
- Human oversight
- Accuracy, robustness and cybersecurity

Data bias

Requirements:
- Risk management systems
- Data and data governance
- Technical documentation
- Record-keeping
- Transparency and provision of information to users
- Human oversight
- Accuracy, robustness and cybersecurity

Data bias

Requirements:

- Risk management systems
- Data and data governance
- Technical documentation
- Record-keeping
- Transparency and provision of information to users
- Human oversight
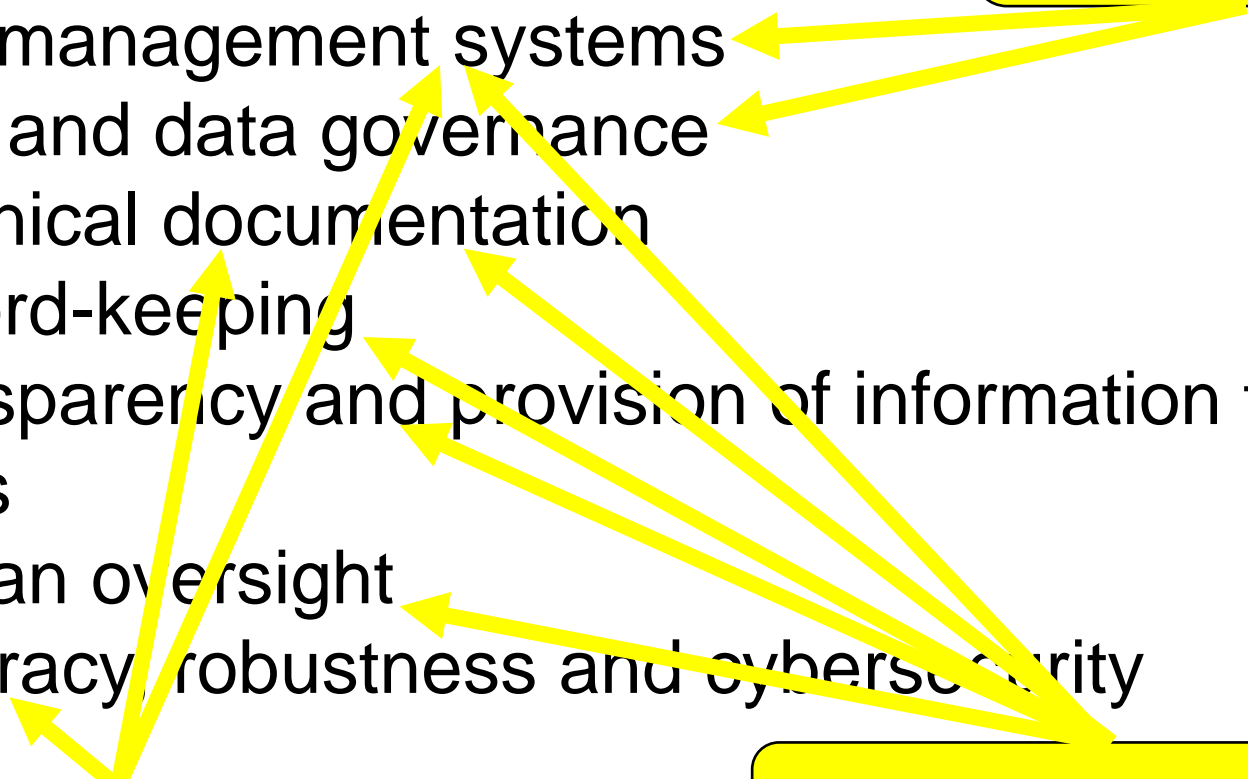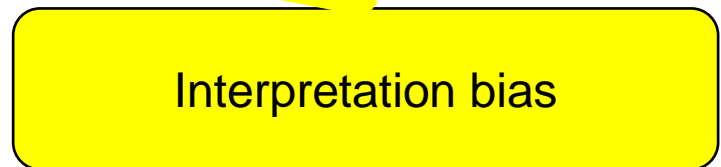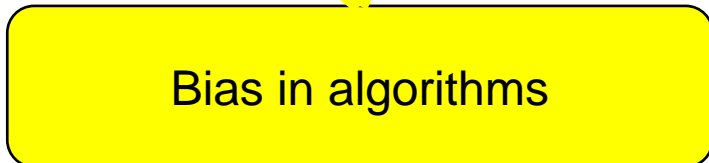- Accuracy, robustness and cybersecurity

Bias in algorithms

Requirements:
- Risk management systems
- Data and data governance
- Technical documentation
- Record-keeping
- Transparency and provision of information to users
- Human oversight
- Accuracy, robustness and cybersecurity

Data bias

Bias in algorithms
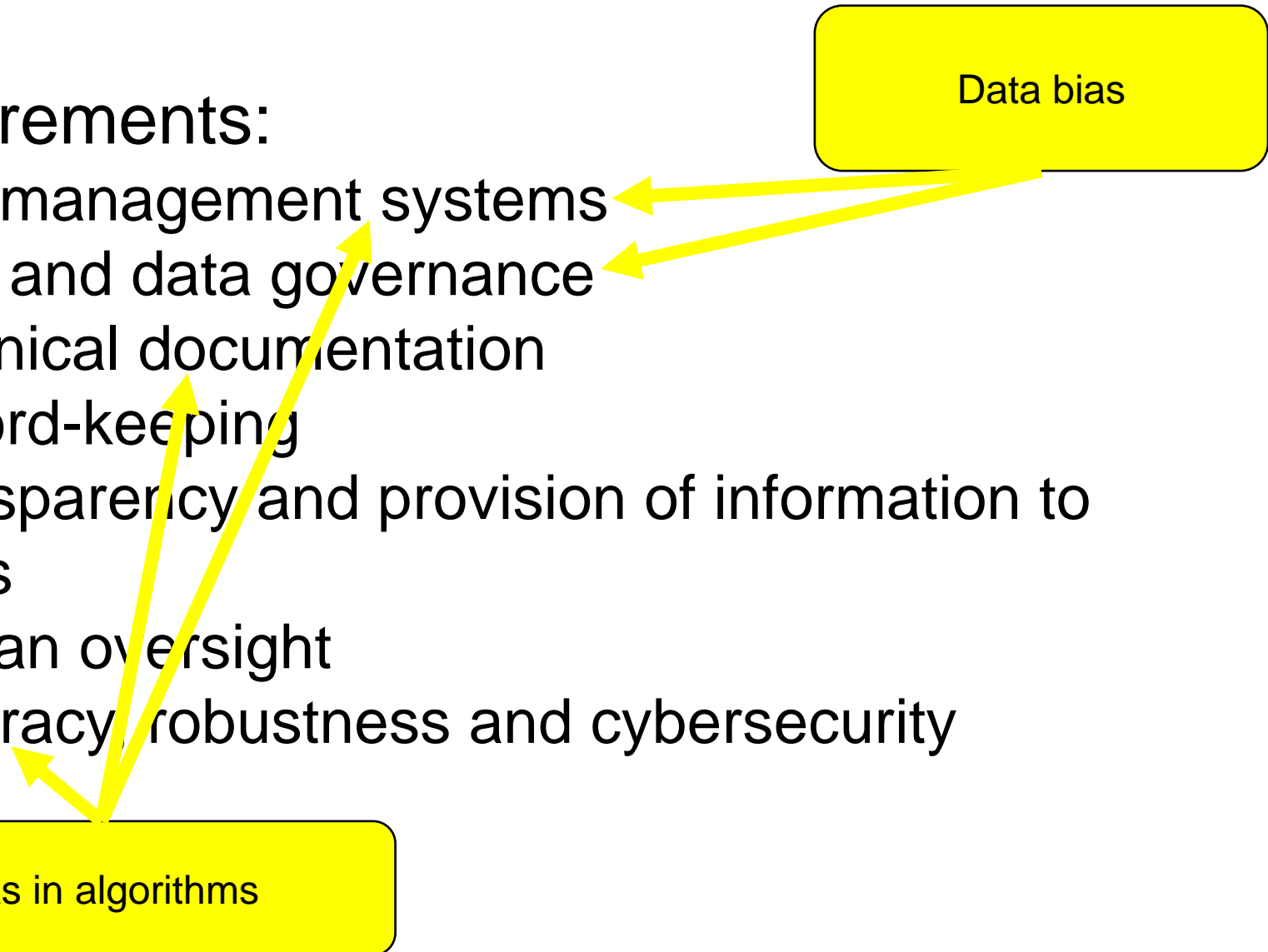
Interpretation bias

Thank you for your attention!

samo.zorc@gov.si