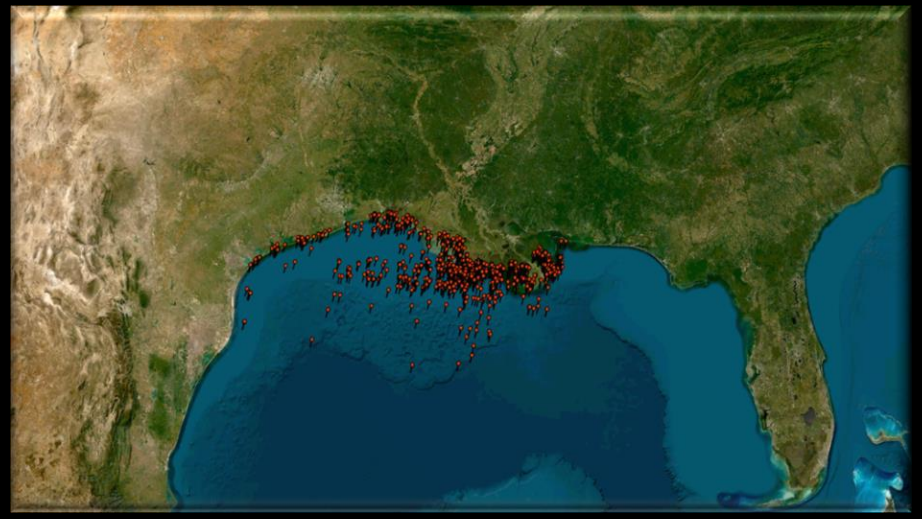
Alexandra Karamitrou
Department of Archaeology
a.Karamitrou@soton.ac.uk

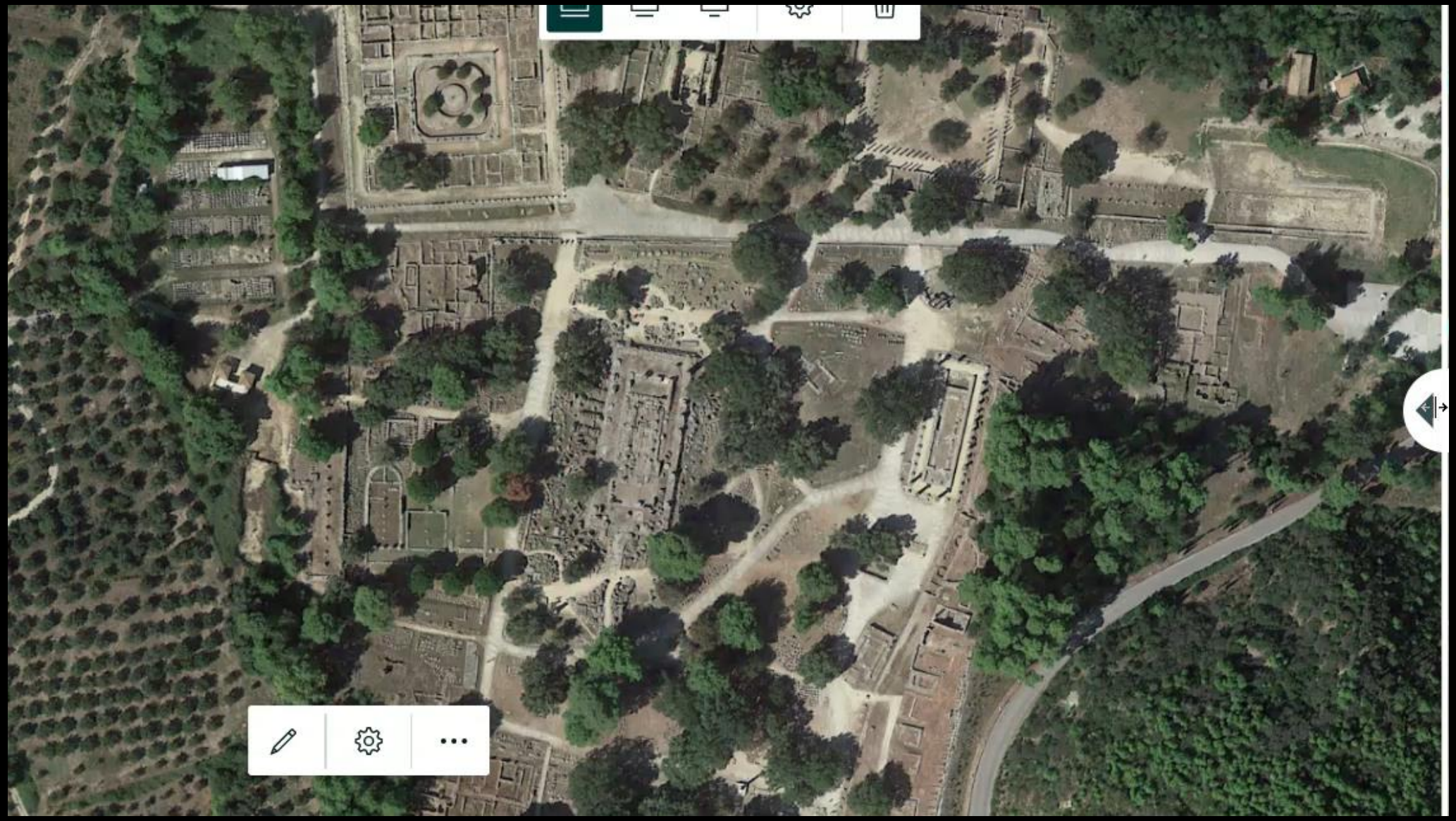
AI Code of Ethics

- Ethical challenges in AI,
- Key values that shape AI systems,
- Actions to ensure they are developed and used responsibly.

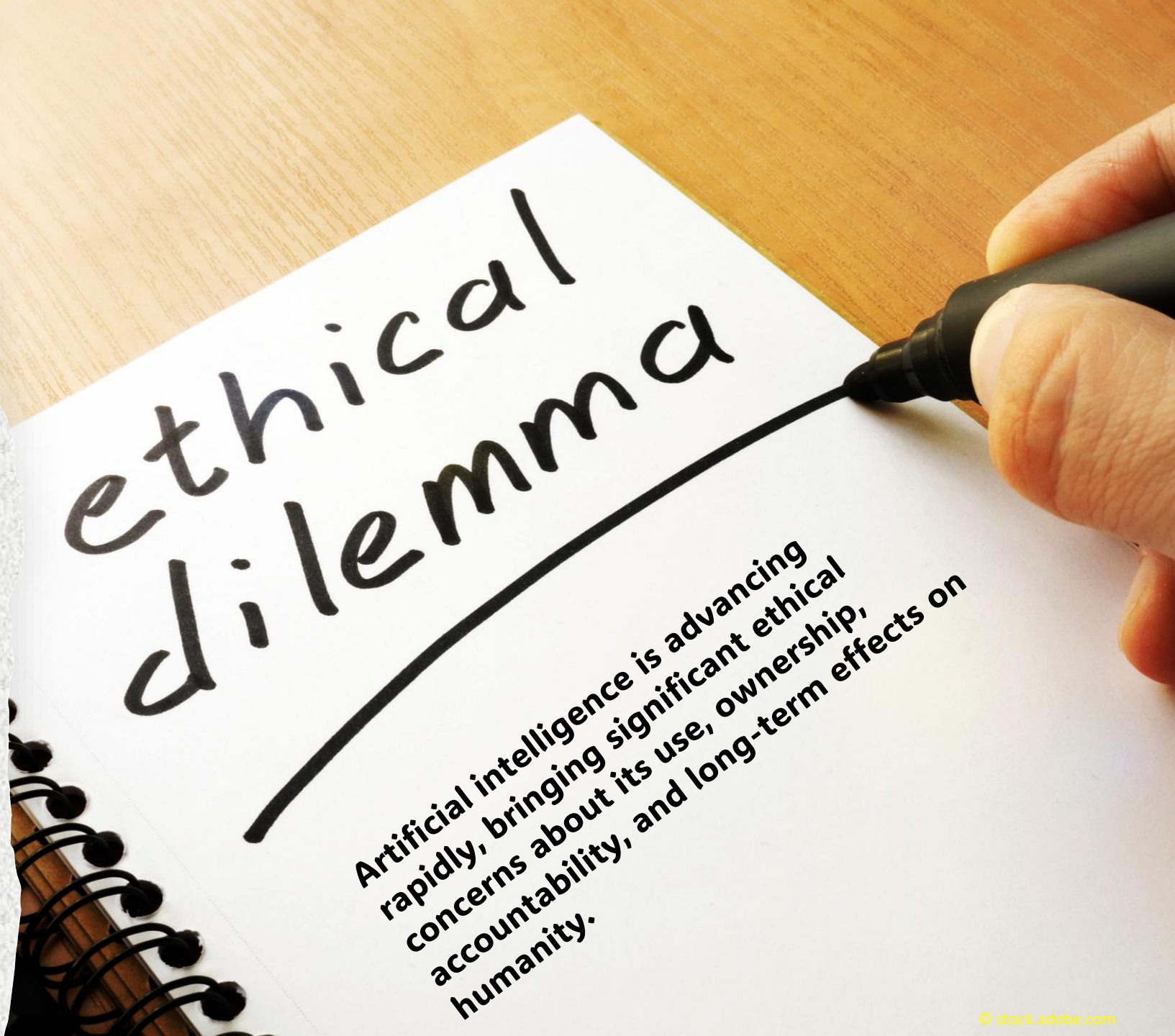
Clear understanding of AI governance

AI in Archaeology




Modern Structure
Vegetation
Background
Archaeological Feature

Like all powerful tools, their development and deployment require great care and responsibility.



Most pressing ethical issues surrounding AI today

- ✓ **Bias and Discrimination**
- ✓ **Transparency and Accountability**
- ✓ **Creativity and Ownership**
- ✓ **Social Manipulation and Misinformation**
- ✓ **Privacy, Security, and Surveillance**

(https://www.captechu.edu/blog/ethical-considerations-of-artificial-intelligence)

Bias and Discrimination

Generating models based on social media data or survey responses can introduce or reinforce biases by excluding already marginalized groups for example. Shaping policies on models trained on such data would introduce these societal inequalities into systems of governance.

Amazon's algorithm discriminated against women

Amazon's automated recruitment system, which was intended to evaluate applicants based on their suitability for various roles. The system learned how to judge if someone was suitable for a role by looking at resumes from previous candidates. As women are still underrepresented in roles relating to STEM, it became biased against women in the process.

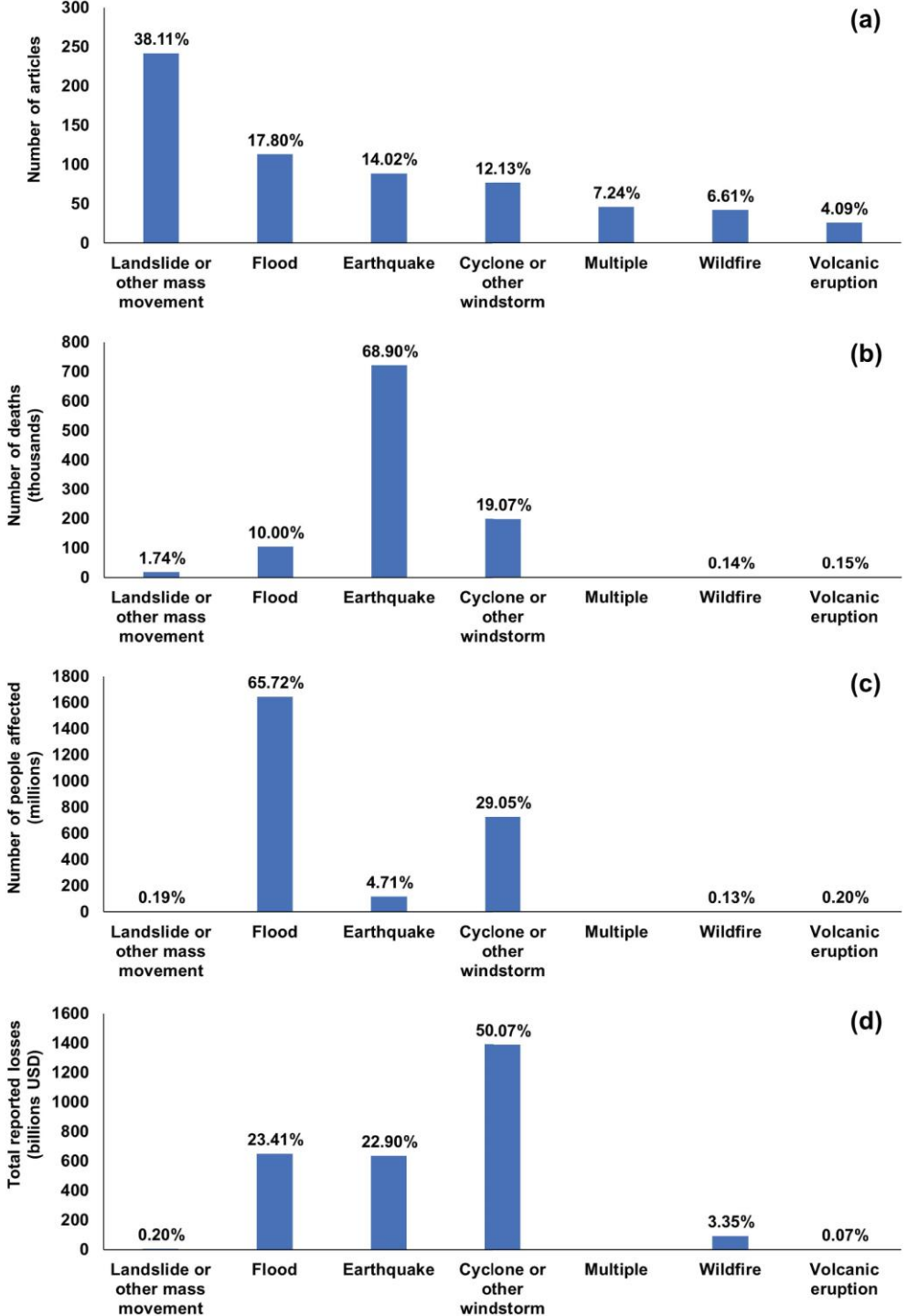
AI in Disaster Risk Management

Remote sensing satellite image data is an important tool and technology that is used to monitor and manage disasters to produce strategic planning models and to predict and control natural disasters as they occur.

Records generated by mobile phones may be used to estimate population sizes before and after a disaster, but may underestimate vulnerable populations who have no access to cell phones.

Humanitarian mapping efforts seem to focus on areas of past disasters, areas containing local mapping communities, and areas of interest for specific stakeholders, such as development agencies. Little information is available for rural or unprioritized areas.

Lack of up-to-date (geospatial) data in low- to middle-income countries compared with high-income countries.



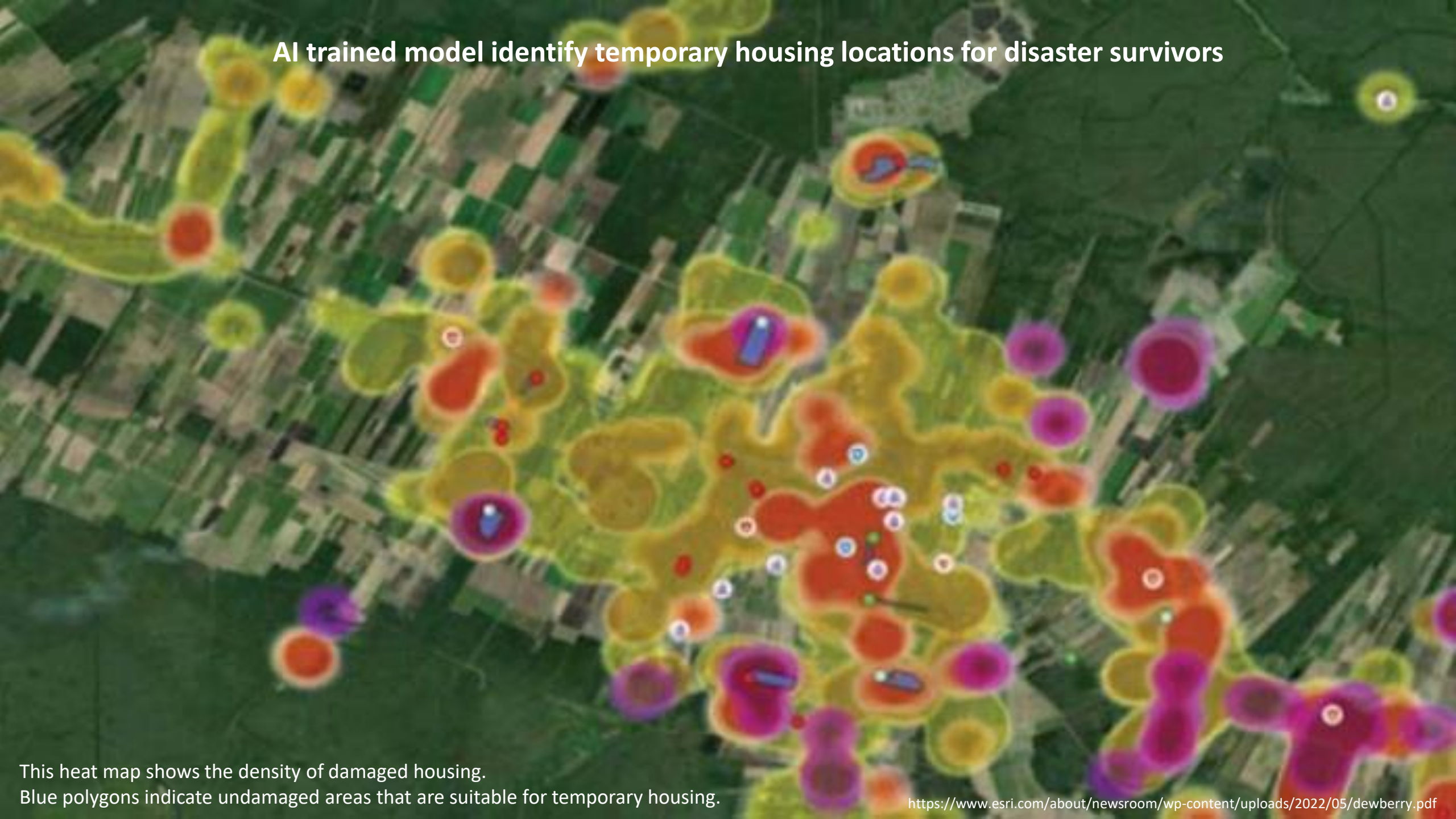
Most studies took place in high-income countries and territories (64%)

The bias towards landslides and other mass movements may likely relate to their relatively small geographic footprint when compared to damages from other hazard types.

Earthquakes, floods, and cyclones and other windstorms are associated with the highest number of deaths, people affected, and economic losses, respectively, compared to the other hazard types.

https://www.sciencedirect.com/science/article/pii/S0034425721002972

AI trained model identify temporary housing locations for disaster survivors



This heat map shows the density of damaged housing.
Blue polygons indicate undamaged areas that are suitable for temporary housing.

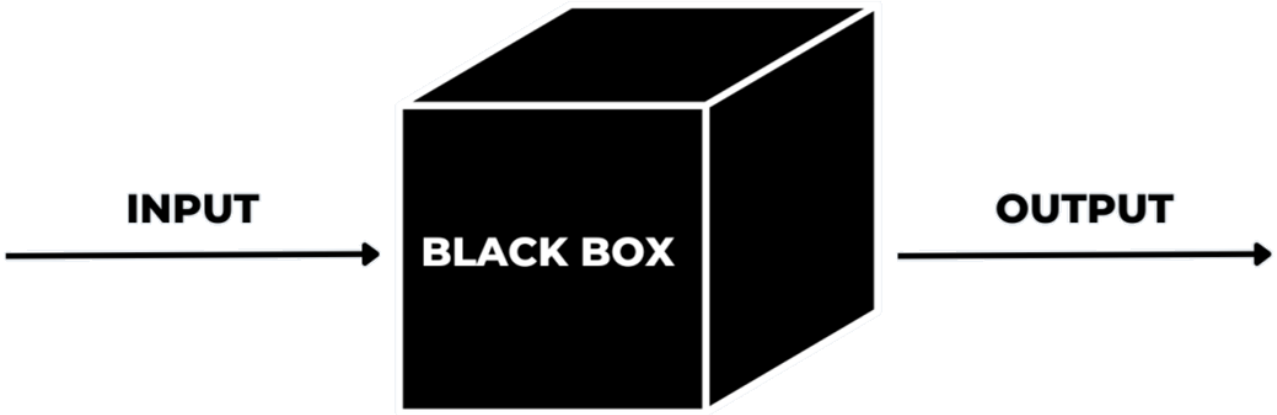
https://www.esri.com/about/newsroom/wp-content/uploads/2022/05/dewberry.pdf

Mumbai, India - where over 60% of the city population lives in slums

Algorithms are not always able to recognize slum buildings as settlements.

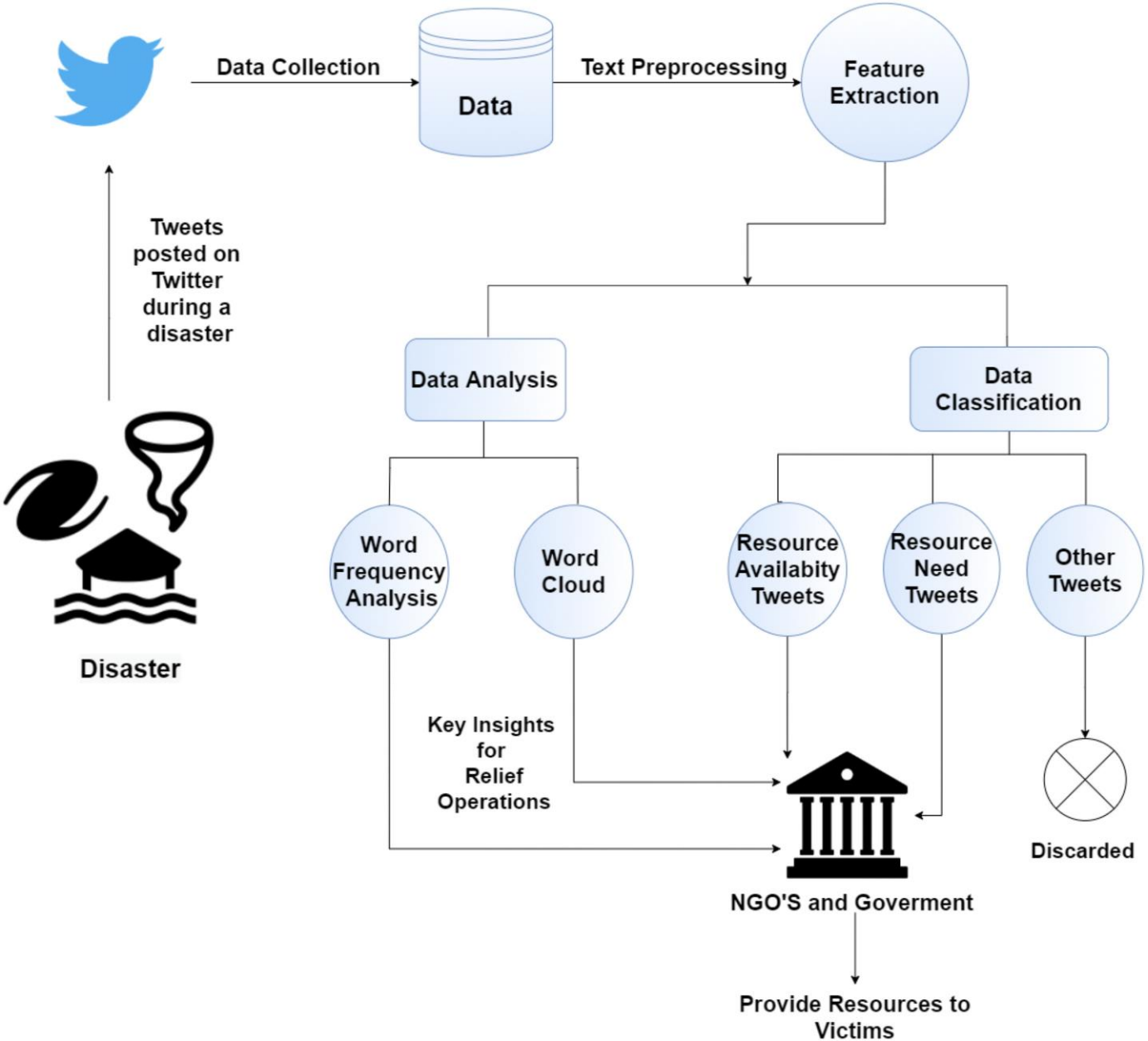
Interpreting remote sensing imagery is context dependent.

Transparency and Accountability



AI systems often function as "black boxes," providing limited transparency on their decision-making processes. it is crucial to understand how these decisions are made and who is responsible for them. Transparency becomes particularly important when AI systems make mistakes or cause harm, as it ensures that appropriate corrective actions can be taken.

Transparency and Accountability

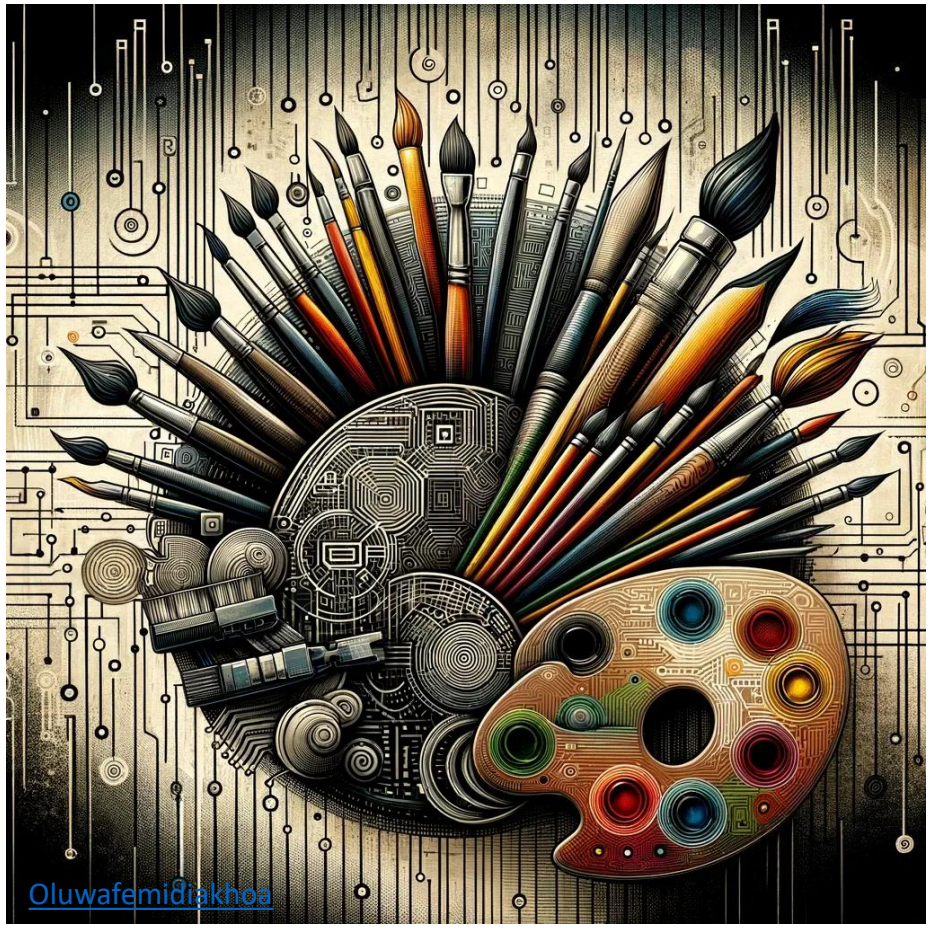


Behl et al., 2021 used explainable AI (Local Interpretable Model-Agnostic Explanations (LIME)) to explore the possible limitations of an algorithm designed to analyze **Twitter** data for identifying people's needs following a disaster.

The results showed that the explanations generated by explainable AI often differ significantly from the way humans usually construct explanations.

Shivam Behl, Aman Rao, Sahil Aggarwal, Sakshi Chadha, H.S. Pannu, Twitter for disaster relief through sentiment analysis for COVID-19 and natural hazard crises, International Journal of Disaster Risk Reduction, Volume 55, 2021, 102101, ISSN 2212-4209, https://doi.org/10.1016/j.ijdrr.2021.102101.

Creativity and Ownership



When a human creates a piece of digital art by entering a text prompt into an AI system that was programmed by a separate individual or organization, questions arise regarding the ownership of the AI-generated art.

- Should we classify AI-generated images as art? What are the qualifications for something to be considered art?
- Who owns the rights to the art?
- Who has the right to commercialize it?
- And who is liable for any potential copyright infringement?

ChatGPT & Dall-E accused of using material without permission

Creators of ChatGPT and the image generation model Dall-E – have been accused of failing to be transparent over what data is used to train their models. This has led to lawsuits from artists and writers claiming that their material was used without permission.

Social Manipulation and Misinformation

AI algorithms can be exploited to spread misinformation, manipulate public opinion, and amplify social divisions.

"deepfakes", "synthetic media", "digitally altered content"

United States, AI-manipulated videos and images of political leaders have made the rounds on social media. Examples include a video that depicted President Biden making transphobic comments and an image of Donald Trump hugging Anthony Fauci.

Shutterstock

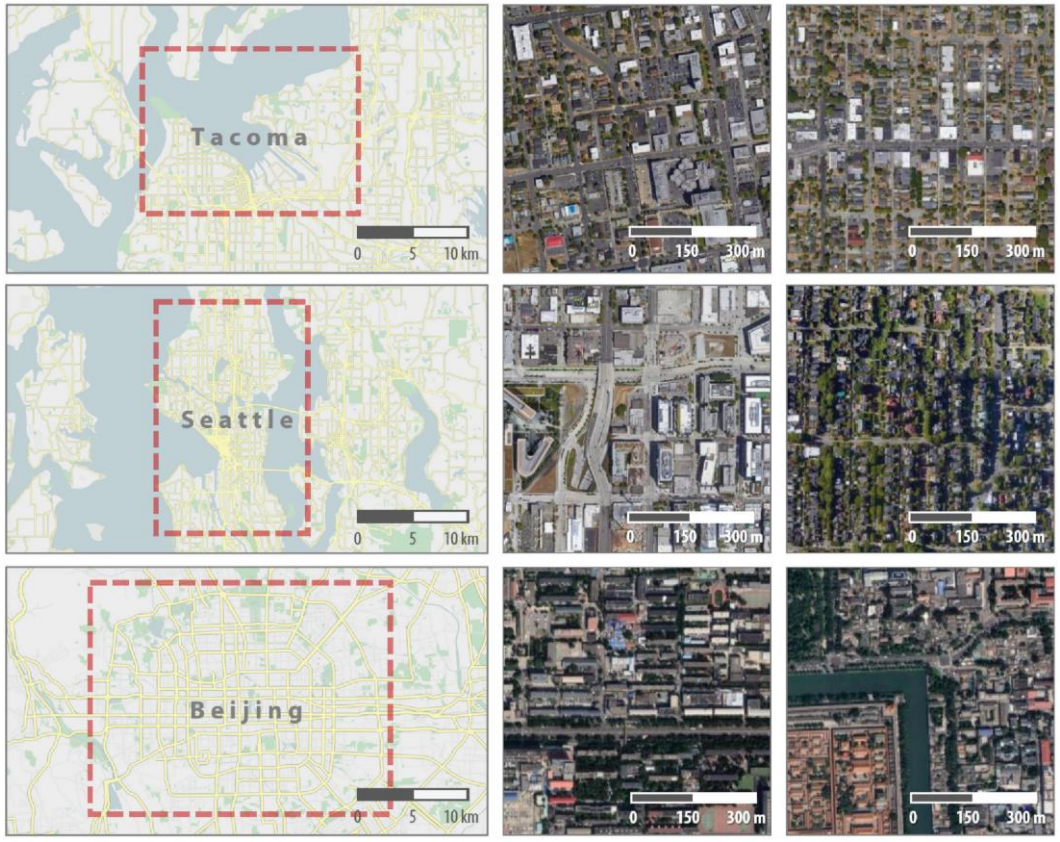
It is estimated that there were million social bots on social media:

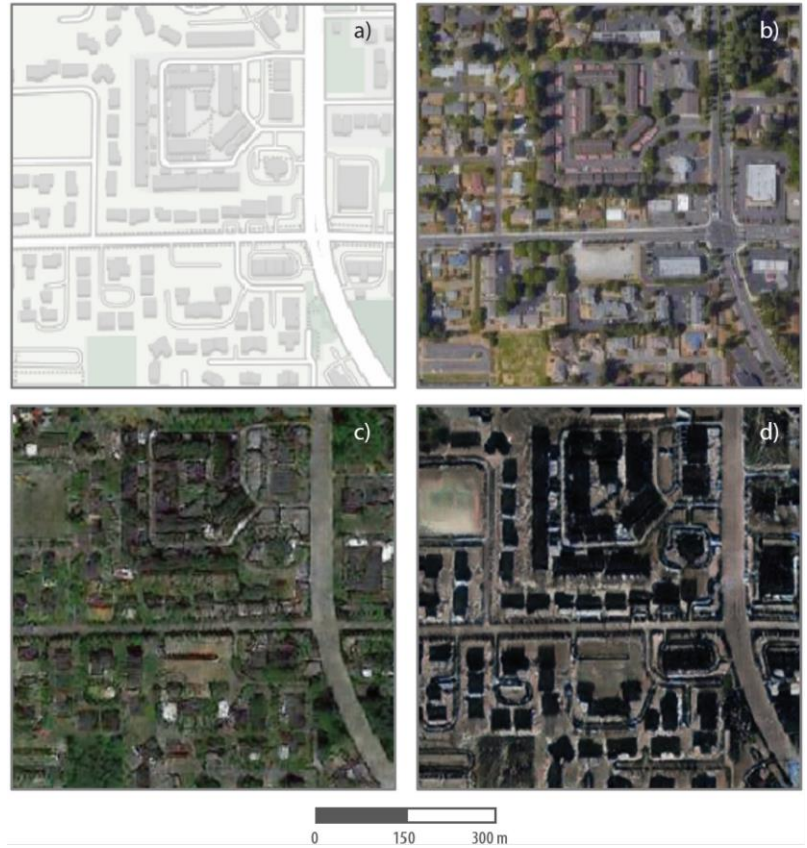
- **Twitter**: 23 million → 8.5% of total users.
- **Facebook**: 140 million → 5.5%–1.2% total users.
- **Instagram**: 27 million → 8.2% of total users

Companies sell fake followers to artificially boost the popularity of accounts.

Social Manipulation and Misinformation

The growing combination of Artificial Intelligence and GIS has sparked concern about the rise of deep fake geography and how it could alter the way people see the geographic world.

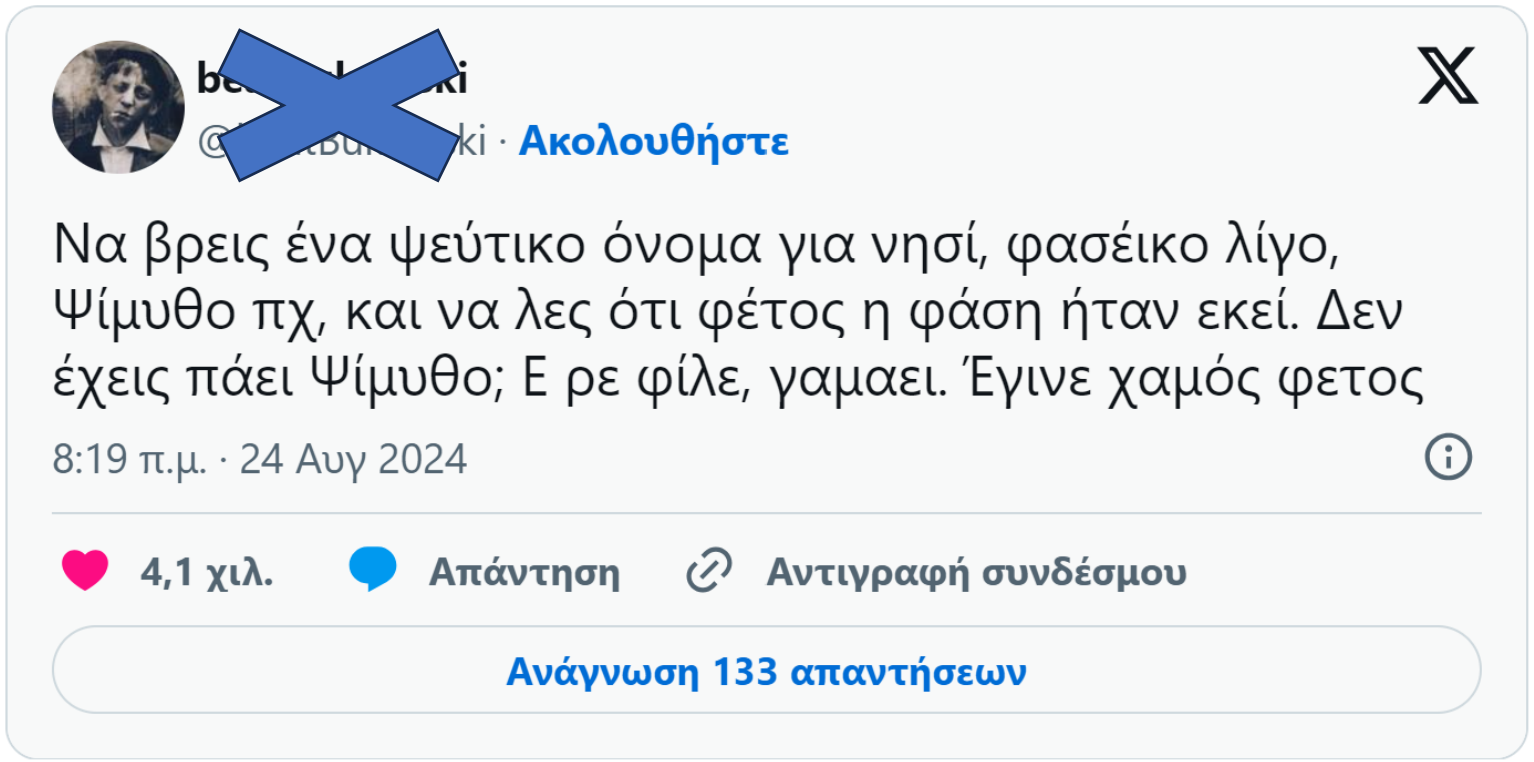




Fake satellite images of a neighborhood in Tacoma with landscape features of other cities. (a) The original basemap tile; (b) the corresponding satellite image tile. The fake satellite image in the visual patterns of (c) Seattle and (d) Beijing.

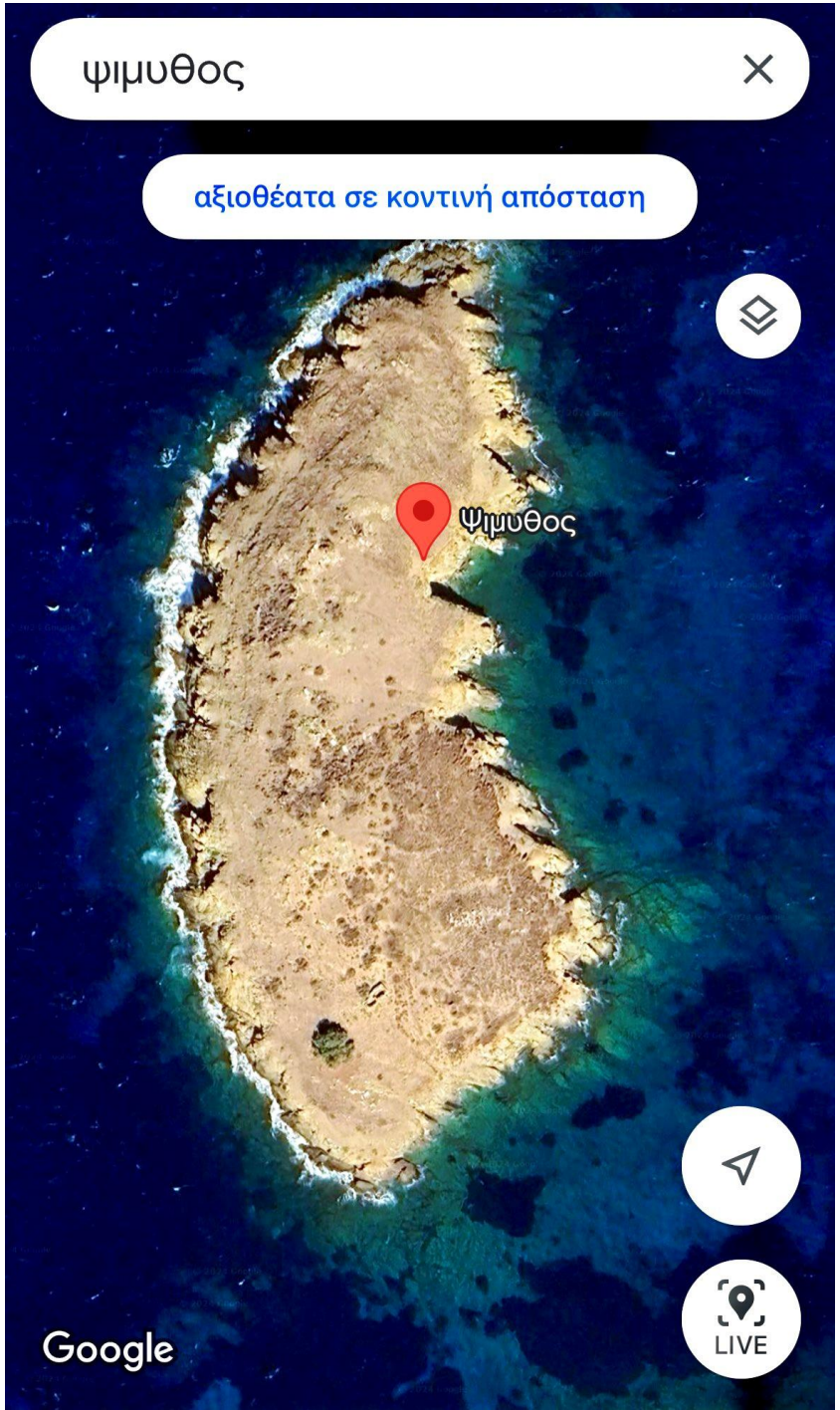
Zhao, B., Zhang, S., Xu, C., Sun, Y., & Deng, C. (2021). Deep fake geography? When geospatial data encounter Artificial Intelligence. *Cartography and Geographic Information Science*, *48*(4), 338–352. https://doi.org/10.1080/15230406.2021.1910075

Fake island in Greece became famous over night!

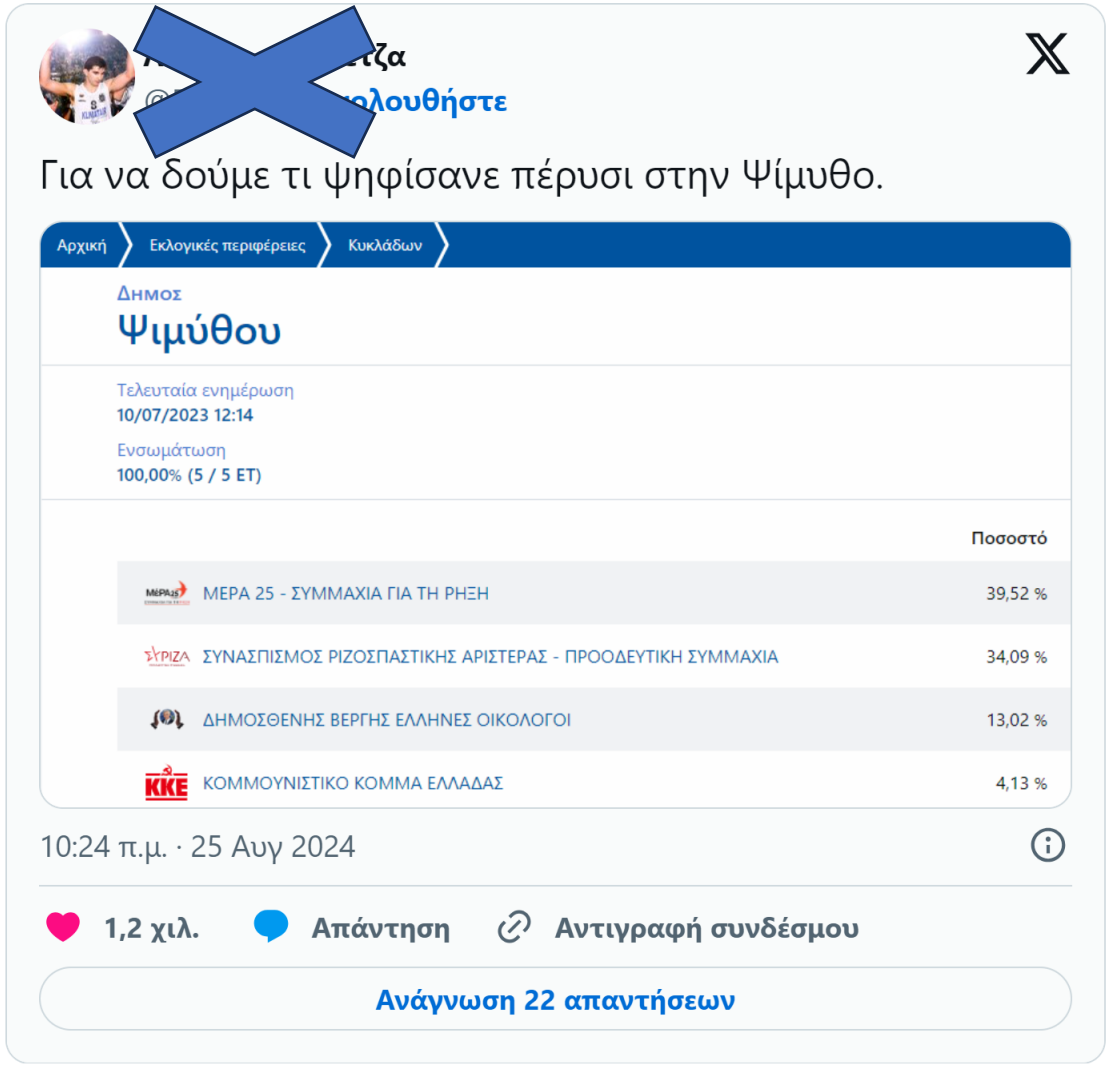


Translation: “To find a fake name for an island, a fake island, Psimythos for example, and say that this year all good events were there. Haven't you been to Psimythos? Hey man, it rules. It was amazing this year”

Fake island in Greece became famous over night!



Fake island in Greece became famous over night!



"Let's see what people of Psithimos voted last year."

TikTok

Search

Log in

For You

Explore New

Following

LIVE

Profile

Log in to follow creators, like videos, and view comments.

Log in

Create TikTok effects, get a reward

Company

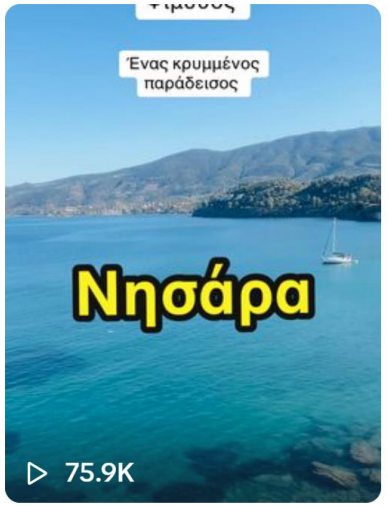
Program

Terms & Policies

Ψιθυμος

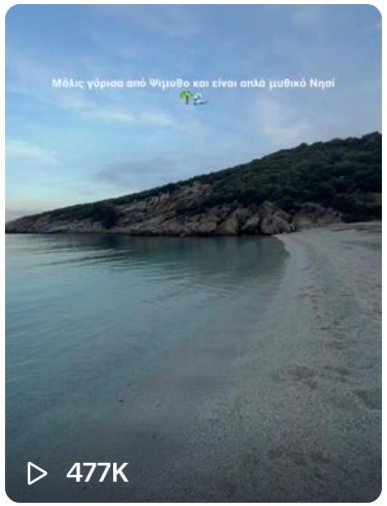
Εσύ το ήξερες? #greektiktok...

[redacted] 9414



🏝️ΨΙΜΥΘΟΣ / τι όμορφο νησι #fy #viral #fypage...

[redacted] 1070

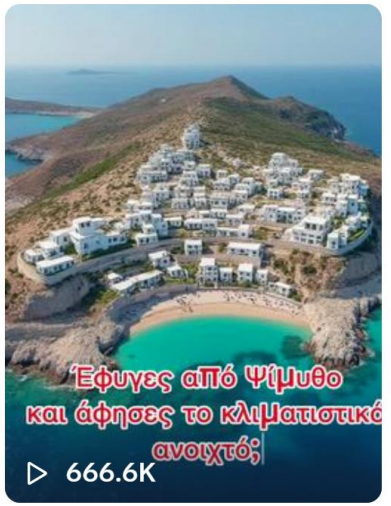


Λενε ότι αν κλείσεις τα μάτια και πεις 'Ψιμυθος'...

[redacted] 14.1K

Πρώτη φορά στην «Ψίμυθο» φέτος και ητα...

[redacted] 67



Πόσοι το πάθατε φέτος το καλοκαίρι; 😜 #inventor...

[redacted] 1426

Privacy, Security, and Surveillance

As AI usage expands, concerns arise regarding how this information is collected, stored, and utilized.

Ascannio/Shutterstock

- How it is being used?
- Who has access to it?

US company “Clearview AI” violated Canadian privacy laws by collecting photographs of Canadian adults and even children for mass surveillance and facial recognition without their consent.

Privacy, Security, and Surveillance

With growing image resolutions and the emergence of drone technologies in Remote Sensing, privacy concerns have become starker and clearer.

While geotagged social media data can aid disaster relief and save lives, its use in assessing housing quality or identifying slum areas risks breaching privacy and stigmatizing individuals, communities, and regions.

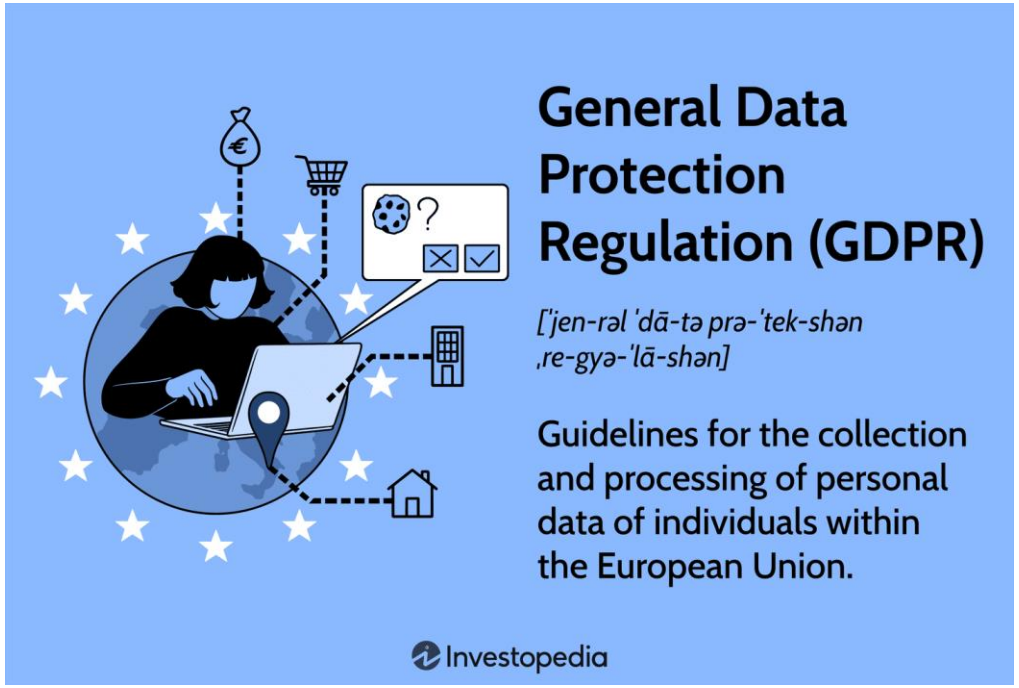
M. Kochupillai, M. Kahl, M. Schmitt, H. Taubenböck and X. X. Zhu, "Earth Observation and Artificial Intelligence: Understanding emerging ethical issues and opportunities," in *IEEE Geoscience and Remote Sensing Magazine*, vol. 10, no. 4, pp. 90-124, Dec. 2022, doi: 10.1109/MGRS.2022.3208357.

Data Governance, Ownership, and Licensing

In the EU, data are protected by laws such as the General Data Protection Regulation (GDPR) and the Directive on Copyright in the Digital Single Market (DSM), making data governance responsible for managing both data privacy and ownership issues.

permit text and data mining for scientific research purposes.

Beyond the GDPR, ethics requires that if data is collected without a person's permission, it must be handled carefully to ensure it doesn't harm their basic rights or well-being.



M. Kochupillai, M. Kahl, M. Schmitt, H. Taubenböck and X. X. Zhu, "Earth Observation and Artificial Intelligence: Understanding emerging ethical issues and opportunities," in *IEEE Geoscience and Remote Sensing Magazine*, vol. 10, no. 4, pp. 90-124, Dec. 2022, doi: 10.1109/MGRS.2022.3208357.

In 2021 UNESCO with its partners developed the Recommendation on the Ethics of Artificial Intelligence, adopted by 193 countries.

unesco

In partnership with

Patrick J McGovern FOUNDATION

The Alan Turing Institute

ITU

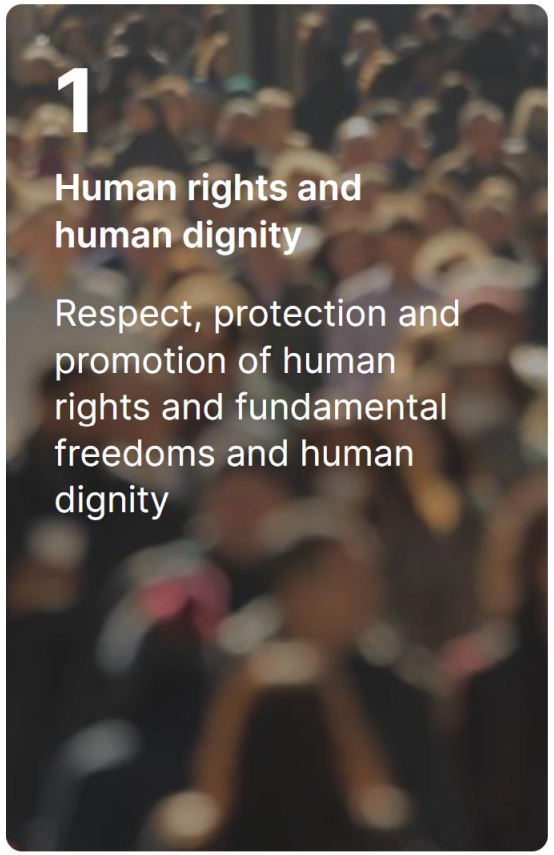
From the People of Japan

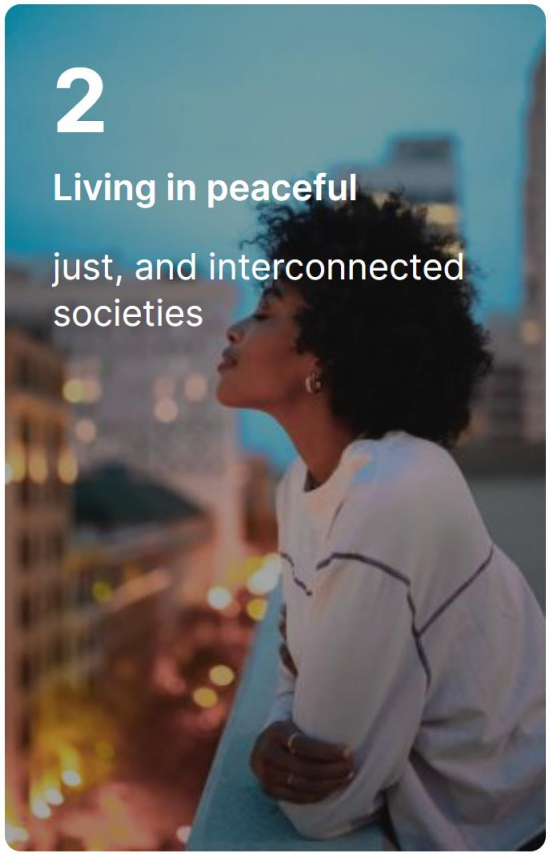
Business Council for Ethics of AI:

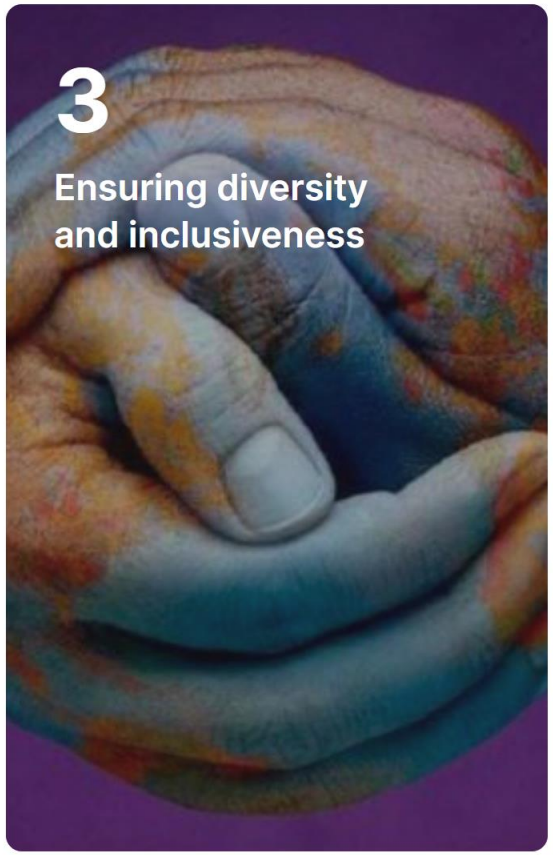
a collaborative initiative between UNESCO & companies involved in the development or use of AI in various sectors.

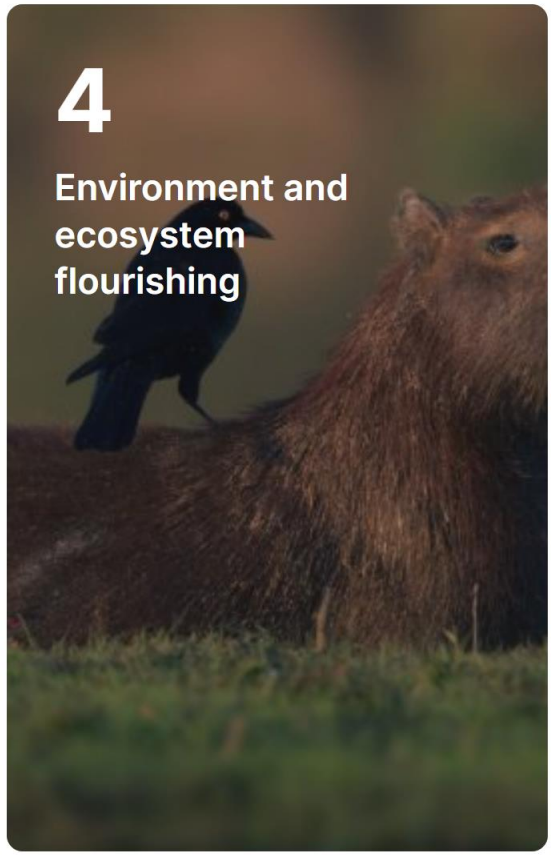
https://unesdoc.unesco.org/ark:/48223/pf0000381137.locale=en

4 core values which lay the foundations for AI systems that work for the good of humanity, individuals, societies and the environment:









https://unesdoc.unesco.org/ark:/48223/pf0000381137.locale=en

Ethical Platform

Help you and your team design and implement AI systems in an ethical, fair, and safe manner, to culture of responsible AI innovation, enabling your team to work together in developing AI technologies that benefit the public good.

- ✓ Ensure that your AI project is ethically permissible by considering the impacts it may have on the wellbeing of affected stakeholders and communities.
- ✓ Ensure that your AI project is fair and non-discriminatory by accounting for its potential to have discriminatory effects on individuals and social groups, by mitigating biases that may influence your model's outputs, and by being aware of the issues surrounding fairness that come into play at every phase of the design and implementation pipeline.
- ✓ Ensure that your AI project is worthy of public trust by guaranteeing to the extent possible the safety, accuracy, reliability, security, and robustness of its product.
- ✓ Ensure that your AI project is justifiable by prioritising both the transparency of the process by which your model is designed and implemented, and the transparency and interpretability of its decisions and behaviours.

https://www.turing.ac.uk/sites/default/files/2019-08/understanding_artificial_intelligence_ethics_and_safety.pdf

Three building-blocks of a responsible AI project delivery ecosystem

Accomplish the goals of establishing the ethical permissibility, fairness, trustworthiness, and justifiability of your project.

Ethical Platform for the Responsible Delivery of an AI Project

SUM Values

that support, underwrite, and motivate a responsible innovation ecosystem

Respect, Connect, Care, Protect

Objectives: to provide an accessible framework for consideration of the moral scope of the social and ethical impacts of your project and to establish well-defined criteria to evaluate its ethical permissibility.

FAST Track Principles

that facilitate an actionable orientation to the ethical design and use of AI systems

Fairness, Accountability, Sustainability, Transparency

Objectives: to make sure that your project is bias-mitigating, non-discriminatory, and fair, and to safeguard public trust in your project's capacity to deliver safe and reliable AI innovation.

PBG Framework

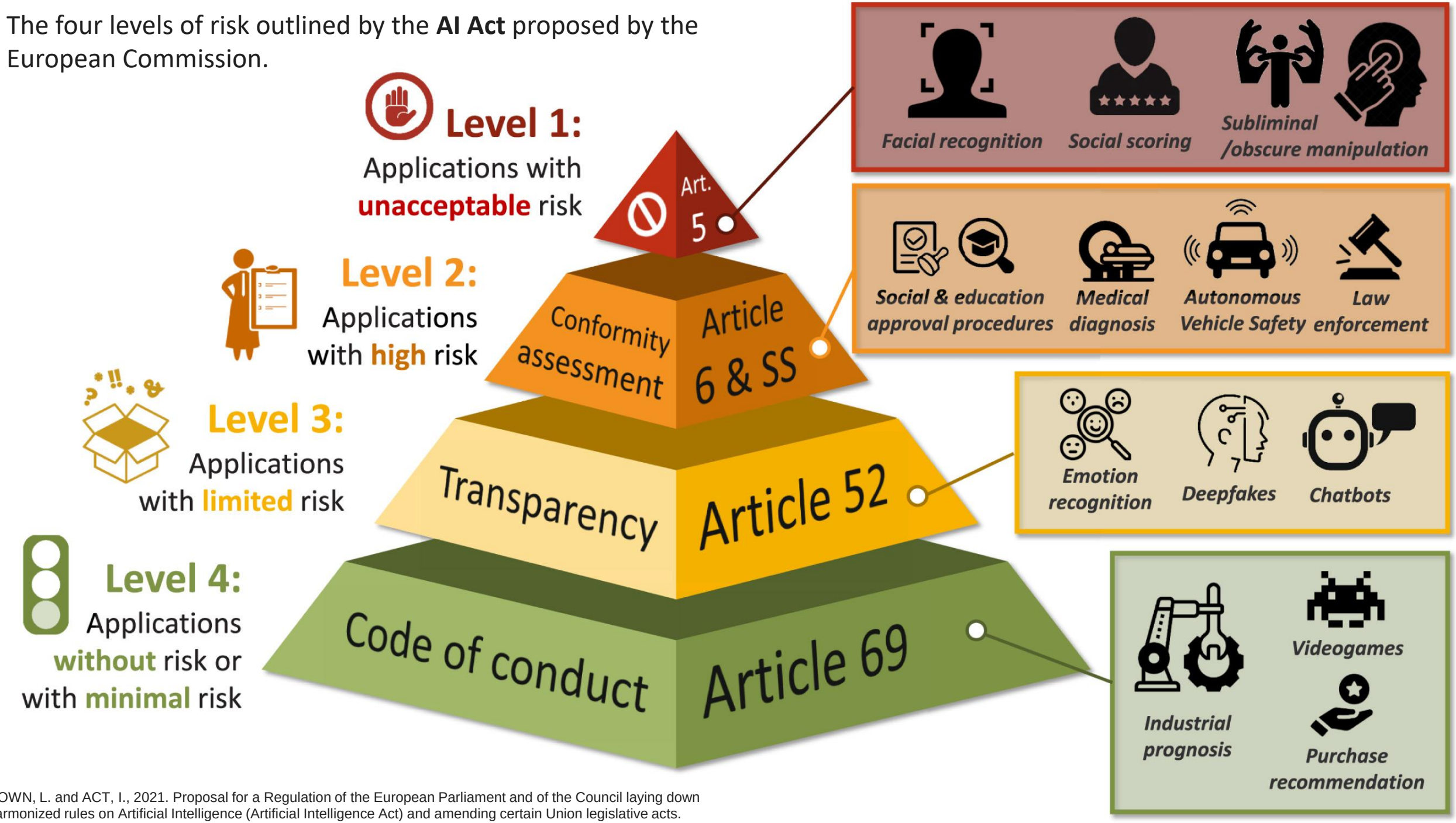
that operationalises the values and principles in an end-to-end workflow governance model

Process-Based Governance Framework

Objective: to set up transparent processes of design and implementation that safeguard the justifiability of both your AI project and its product as well as enable end-to-end accountability.

https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

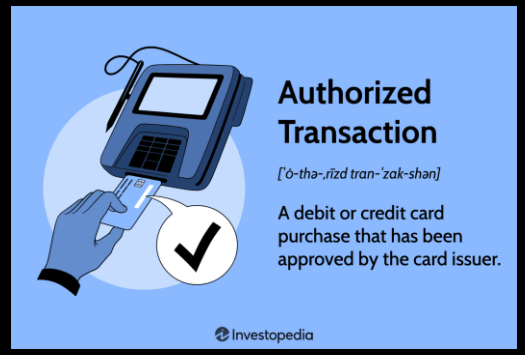
The four levels of risk outlined by the **AI Act** proposed by the European Commission.



DOWN, L. and ACT, I., 2021. Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts.

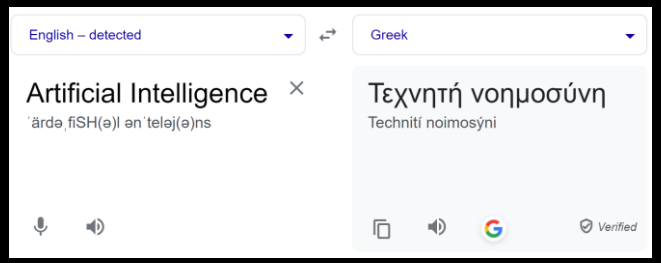
AI in our lives

Credit card approved transaction



GPS

Google Translate Service



Smart Phones

Vacuum Cleaner

Baxter robot

Medicine

Education

Transportation

Economy

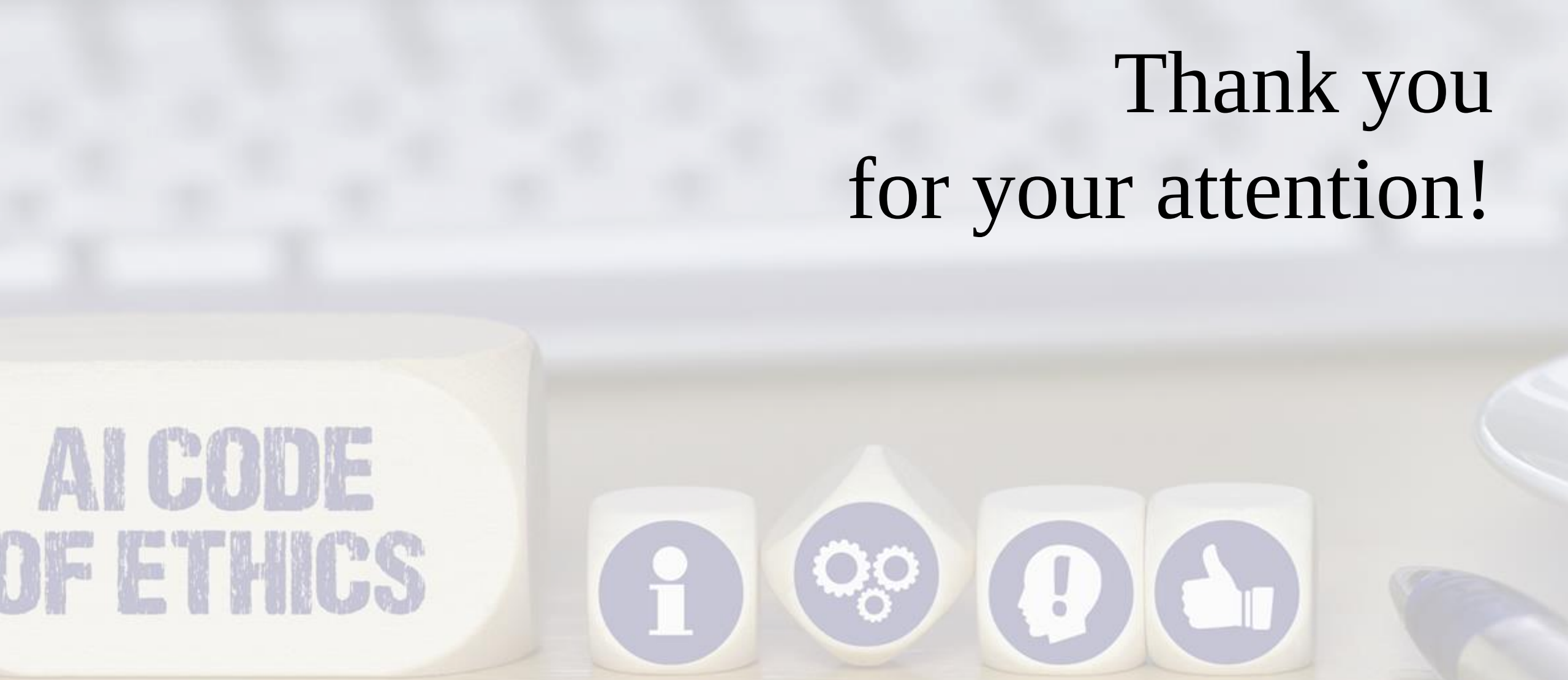
Useful Resources

Recommendation on the Ethics of Artificial Intelligence

Understanding artificial intelligence ethics and safety (PDF)

Proposal for a Regulation laying down harmonised rules on artificial intelligence

Thank you for your attention!



All questions are welcome

Alexandra Karamitrou
a.Karamitrou@soton.ac.uk