Should Artificial Intelligence be regulated?

"…in recent decades, however, a consensus has emerged around the idea of a rational agent that perceives and acts in order to maximally achieves its objectives"….. *"Up to now,* ***AI has focused on systems that are better at making decisions; but this is not the same as making better decisions…..well aligned with human values"****.*

AI should be beneficial with applications related to health, climate change, energy, smart cities, food, equity, inclusion and sustainability at large, but it can also be applied to applied to dangerous applications like the ones on autonomous weapons.

Also, AI will bring substantial societal impacts: job losses, fake news generation, election control through social influence, personal data privacy

EU has delivered guidelines for trustworthy AI.

Stuart Russell, "Provably Beneficial Artificial Intelligence", 2017
https://people.eecs.berkeley.edu/~russell/papers/russell-bbvabook17-pbai.pdf

Etzioni, Amitai, and Oren Etzioni. "Should Artificial Intelligence Be Regulated?" *Issues in Science and Technology* 33, no. 4 (Summer 2017).

Ethics guidelines for trustworthy AI

- Delivered by the European Commission's High-Level Expert Group on Artificial Intelligence (AI HLEG).
- "Trustworthy AI has two components: (1) it should respect fundamental rights, applicable regulation and core principles and values, ensuring an "ethical purpose" and (2) it should be technically robust and reliable since, even with good intentions, a lack of technological mastery can cause unintentional harm."
- "Incorporate the requirements for Trustworthy AI from the earliest design phase: Accountability, Data Governance, Design for all, Governance of AI Autonomy (Human oversight), Non-Discrimination, Respect for Human Autonomy, Respect for Privacy, Robustness, Safety, Transparency."
- "Foresee training and education, and ensure that managers, developers, users and employers are aware of and are trained in Trustworthy AI."

- **https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai**

Which properties of AI?

Fairness: Decisions should not be discriminatory. We should be sure for instance that race or gender are not influencing decisions. But data are biased. Amazon automatic curricula selector was giving preferences to male candidates as the data set was biased (experiment closed in 2017). Microsoft chatbot Tay learning from Twitts to behave as a nazist.

Transparency: A system behaviour should be understandable under every circumstances and based on a comprehensible model.

Verifiability: Formally prove that the system is correct with respect to some property.

Explainability: Being able to explain the decision taken and the factors that have determined it.

Accountability: responsibility for the decision taken.

Accuracy, Privacy …..

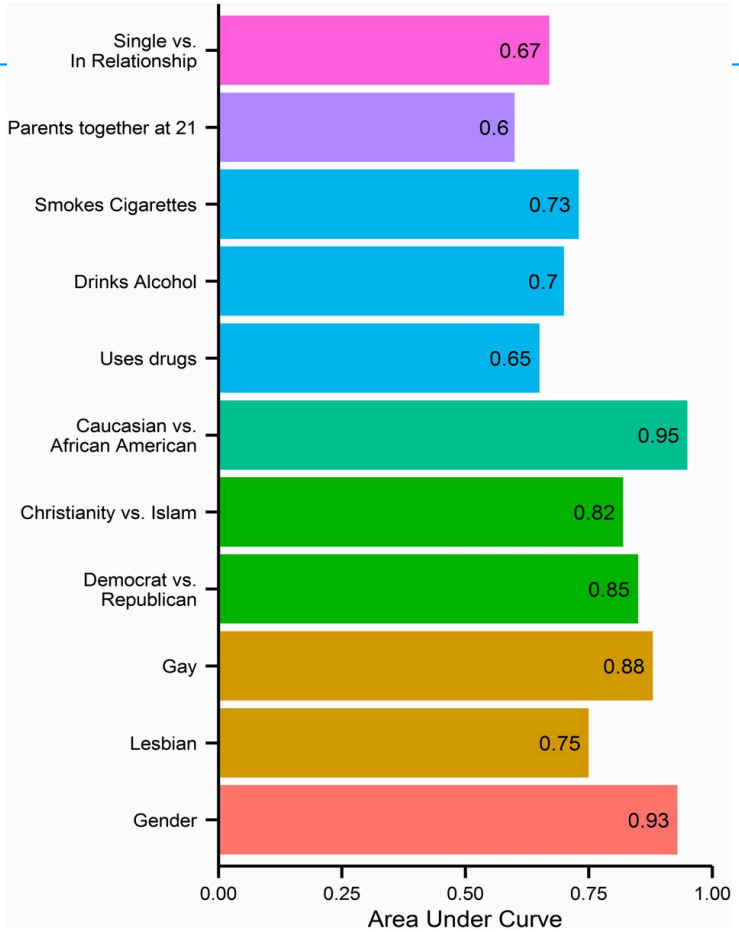
These features are not always all needed. For instance I would like to understand why I am not eligible for a loan, maybe I am not interested in understanding why the hoover made a given path.

Informatics Europe e ACM Europe Council
When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making ACM, "Statement on Algorithmic Transparency and Accountability," 12 January 2017.

"Private traits are predictable from digital records of human behavior"

"Easily accessible digital records of behavior, Facebook Likes, can be used to ***automatically and accurately predict a range of highly sensitive personal attributes****: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender"*

- A dataset of 58,000 volunteers have made available their Facebook Likes and detailed personal data, profiles etc. for machine learning
- The learnt model is enough accurate and discriminates among different categories (homosexual and eterosexuality with 88% accuracy, Democratics and Repubblicans with 85% accuracy).
- Implications on privacy (Cambridge Analytica).



Michal Kosinski, David Stillwell and Thore Graepel. "Private traits and attributes are predictable from digital records of human behavior". PNAS April 9, 2013. 110 (15) 5802-5805

Explainable AI is an open challenge

- It is not easy to make a sub-symbolic system explainable. The AI community is actively working on these themes.
- Not yet general ideas, but good promising results on specific domains.
- Fairness, Interpretability, Explainability ECAI-IJCAI Workshops 2018.

- Toward **an integration of the two souls of AI** for combining the advantages of both approaches in hybrid architectures.

- Integrate **deep learning**, which is excellent for perception and machine learning (but is **a black box**) with **symbolic systems** that are **trasparent** and are able to perform reasoning and abstraction.

"Learning Explanatory Rules from Noisy Data", Richard Evans, Edward Grefenstette DeepMind, London, UK Journal of Artificial Intelligence Research (2018).
"Neural-Symbolic Learning and Reasoning: Contributions and Challenges" Artur d'Avila Garcez et alii., The 2015 AAAI Spring Symp., 2015.
L.G. Valiant, "Knowledge Infusion: In Pursuit of Robustness in Artificial Intelligence". In FSTTCS 2008.
Probabilistic Inductive Logic Programming Editors: De Raedt, L., Frasconi, P., Kersting, K., Muggleton, S.H. LNCS 4911 2008.

Human-centered Artificial Intelligence

The more successfull is a technology the more invisible it is

European statistics claim that in 2014 25.900 people died. And four times higher is the number of people severly and permanently njured.

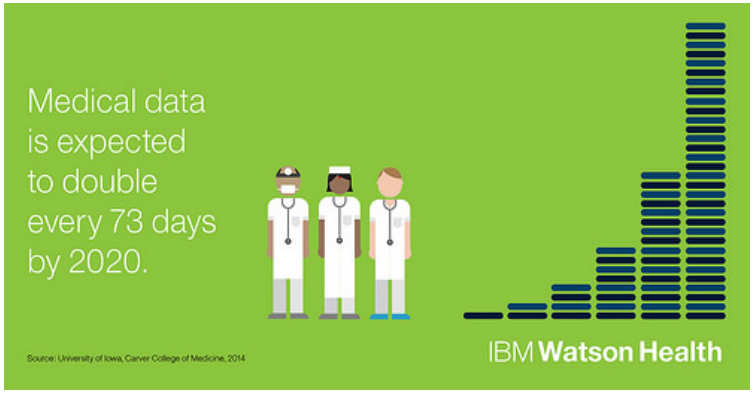
ASIANSCIENTIST

TOP NEWS IN THE LAB HEALTH TECHNOLOGY PHARMA ACADEMIA FEATURES

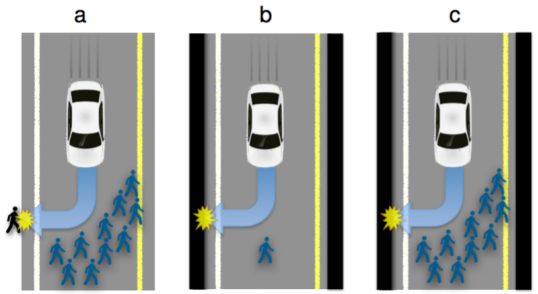
IBM's Watson Detected Rare Leukemia In Just 10 Minutes

The supercomputer swiftly cross-referenced a patient's genetic data to make a diagnosis that would have taken a human doctor weeks.

SHARE

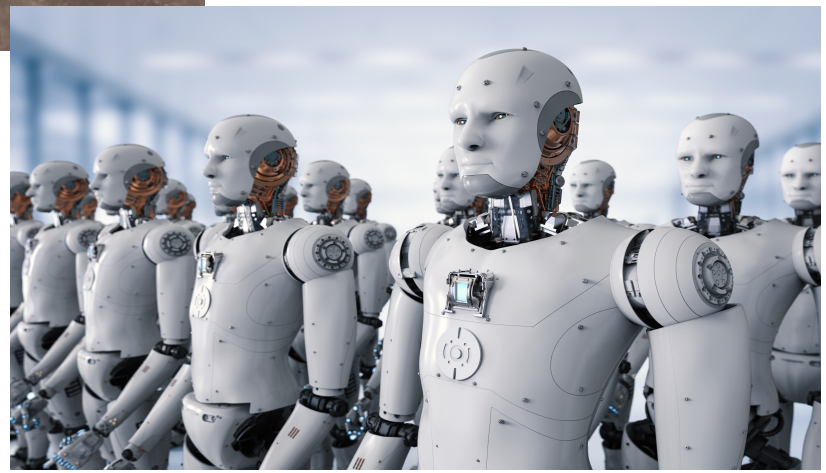


Ethical, Legal and Social aspects



http://moralmachine.mit.edu

Autonomous weapons



McKinsey report claims that in 2030 we will have a 30% of job losses for AI

Toward a beneficial AI

"Up to now, AI has focused on systems that **are better at making decisions**; but this is **not the same as making better decisions**. No matter how excellently an algorithm maximizes, and no matter how accurate its model of the world, a machine's decisions may be ineffably stupid, in the eyes of an ordinary human, **if its utility function is not well aligned with human values**. …**This problem requires a change in the definition of AI itself, from a field concerned with pure intelligence, independent of the objective, to a field concerned with systems that are provably beneficial for humans.**"

("Provably Beneficial Artificial Intelligence" , Stuart Russell 2017).

Beneficial AI

"….a change in the definition of AI itself, from a field concerned with pure intelligence, independent of the objective, to a field concerned with systems that are provably beneficial for humans."

("Provably Beneficial Artificial Intelligence" , Stuart Russell 2017).

- AI can have a beneficial impact on society, economy and environment, the three pilars of sustainable development.
- **Computational Sustainability** (NSF expedition)
- Mixture of computational techniques for dealing with big societal challenges

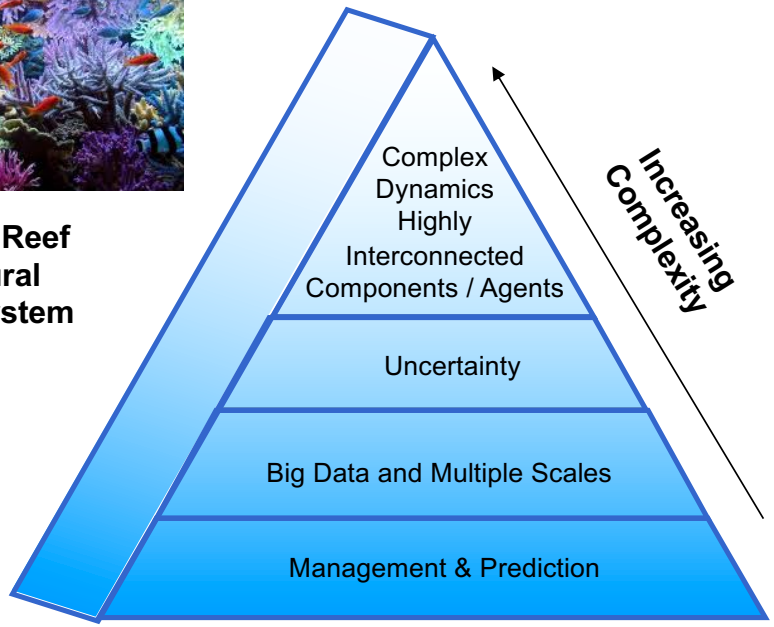
Beneficial AI

Sustainability problems
unique in scale
and complexity

Smart Power Grid:
Complex Digital Ecosystem

Smart, connected and resilient CIties

Coral Reef
Natural
Ecosystem



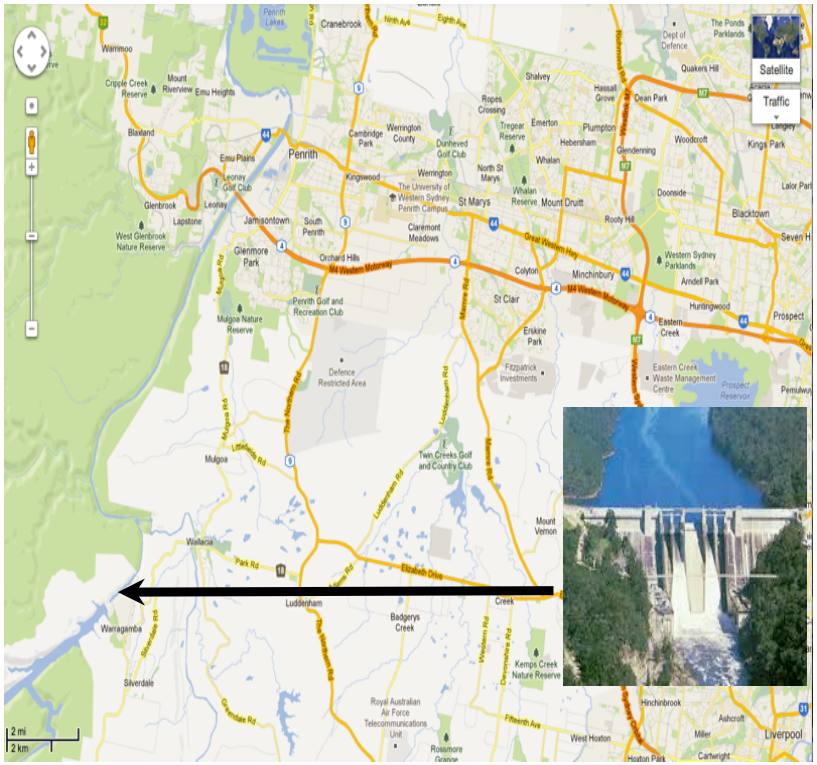
Significant Computational Challenges

Flood management

People involved
- 70,000 persons

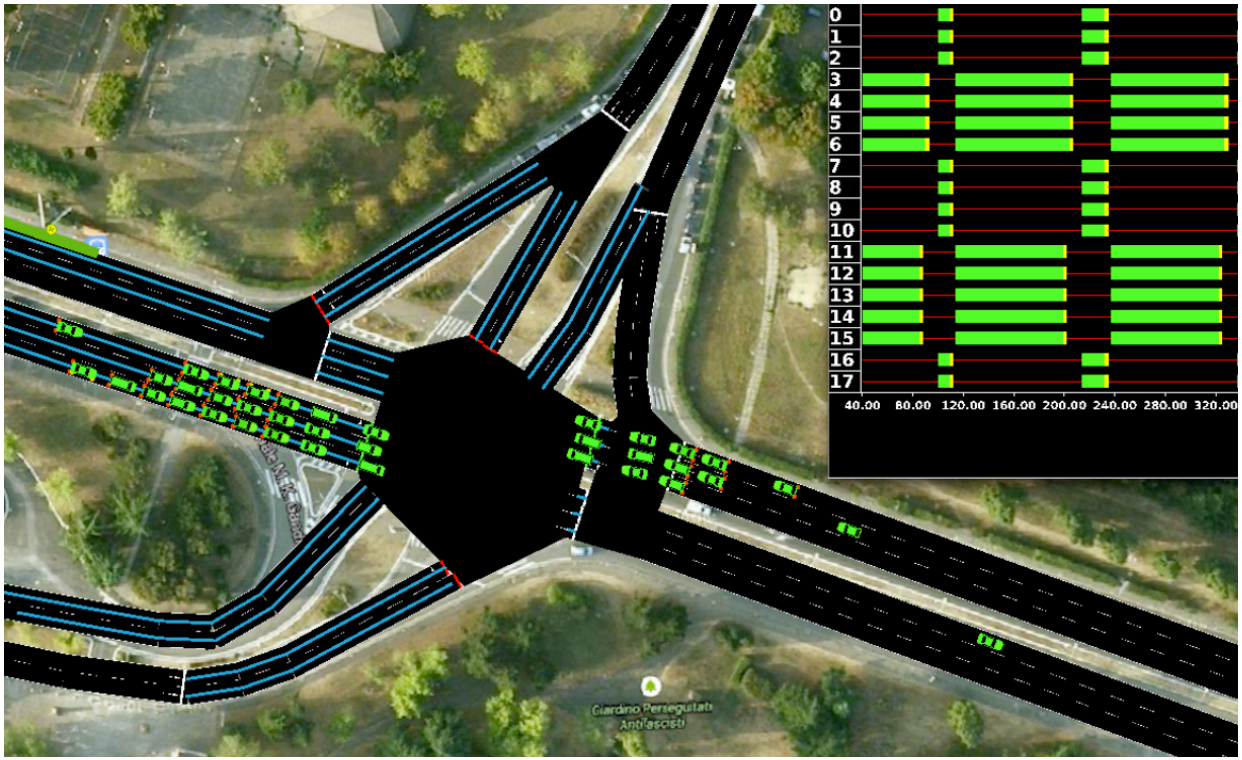
Evacuation profile
- 50 residential zones (evacuation nodes)
- 10 evacuation centers (safe nodes)
- 125 transit nodes (intersections)
- 458 edges (road segments)

Flooding scenarios
- Past data on previous floods



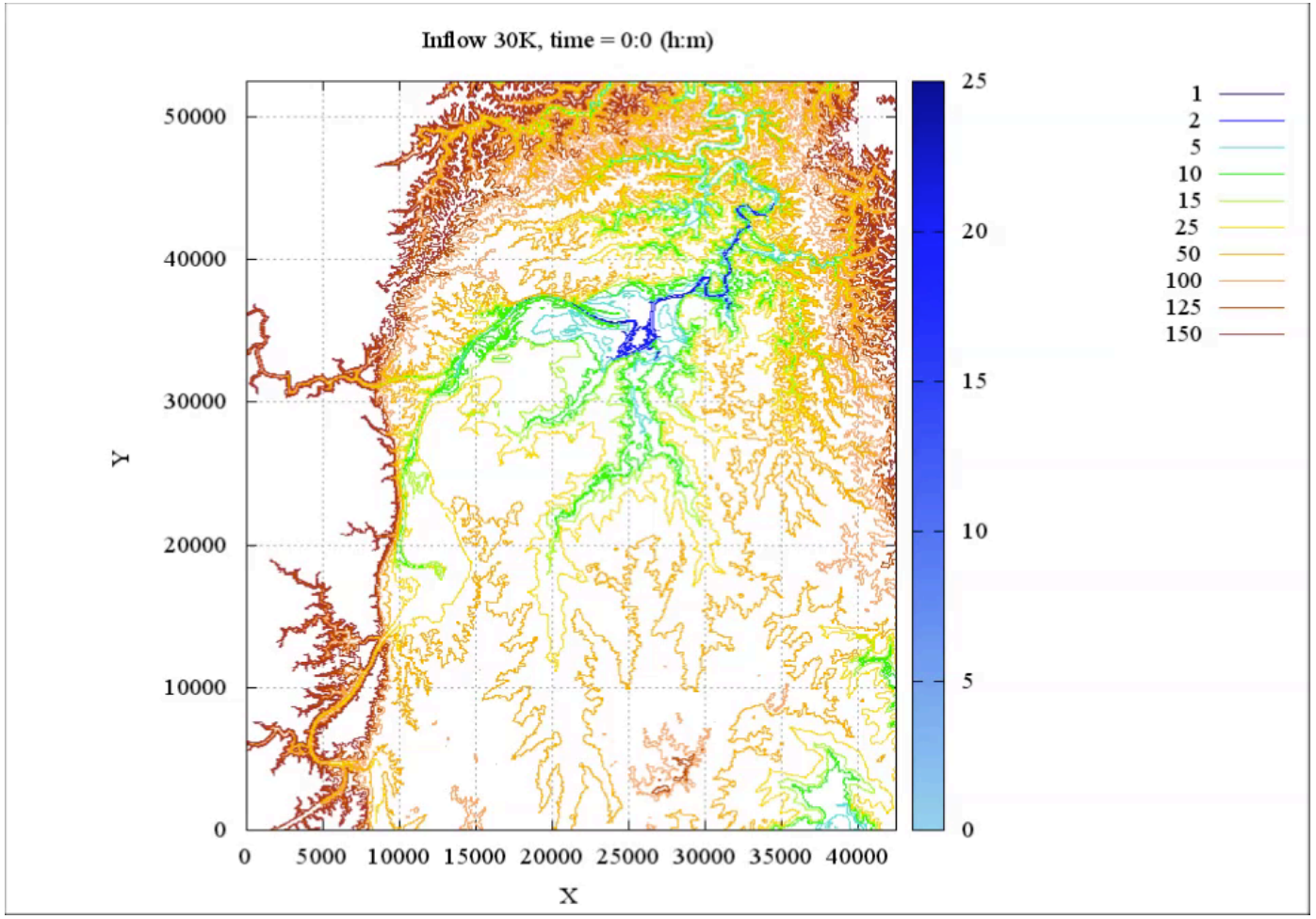
INTEGRATION OF DESCRIPTIVE – PREDICTIVE – PRESCRIPTIVE MODELS

Descriptive model: road network and infrastructure



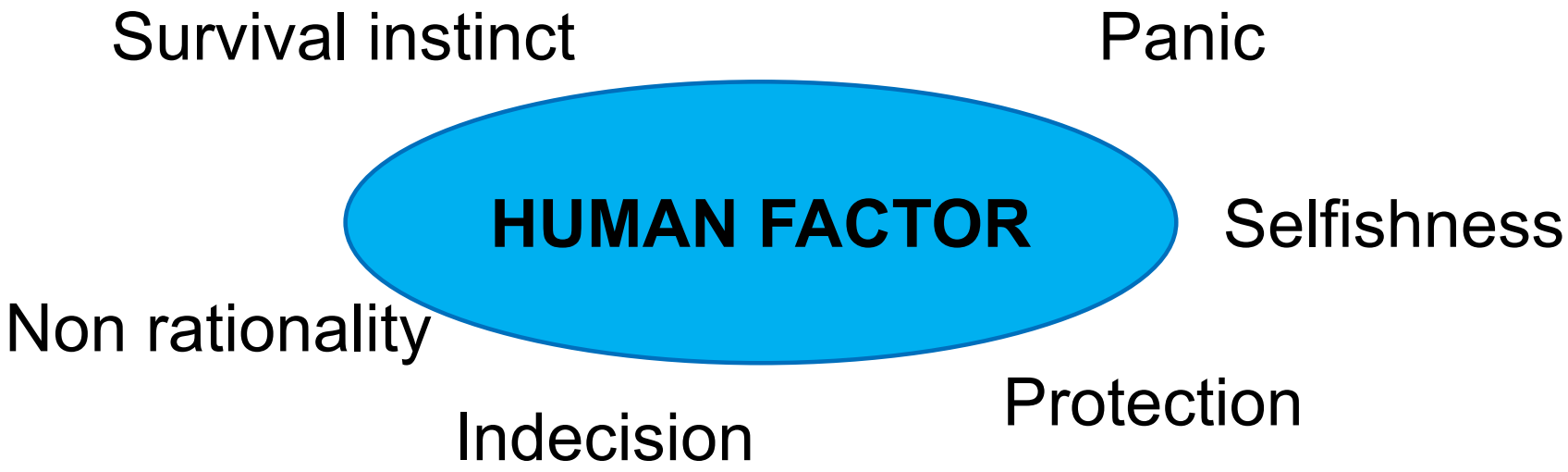
Road network and infrastructure
Max flow for each road
Population density
Past data

Flood Simulation: predictive model



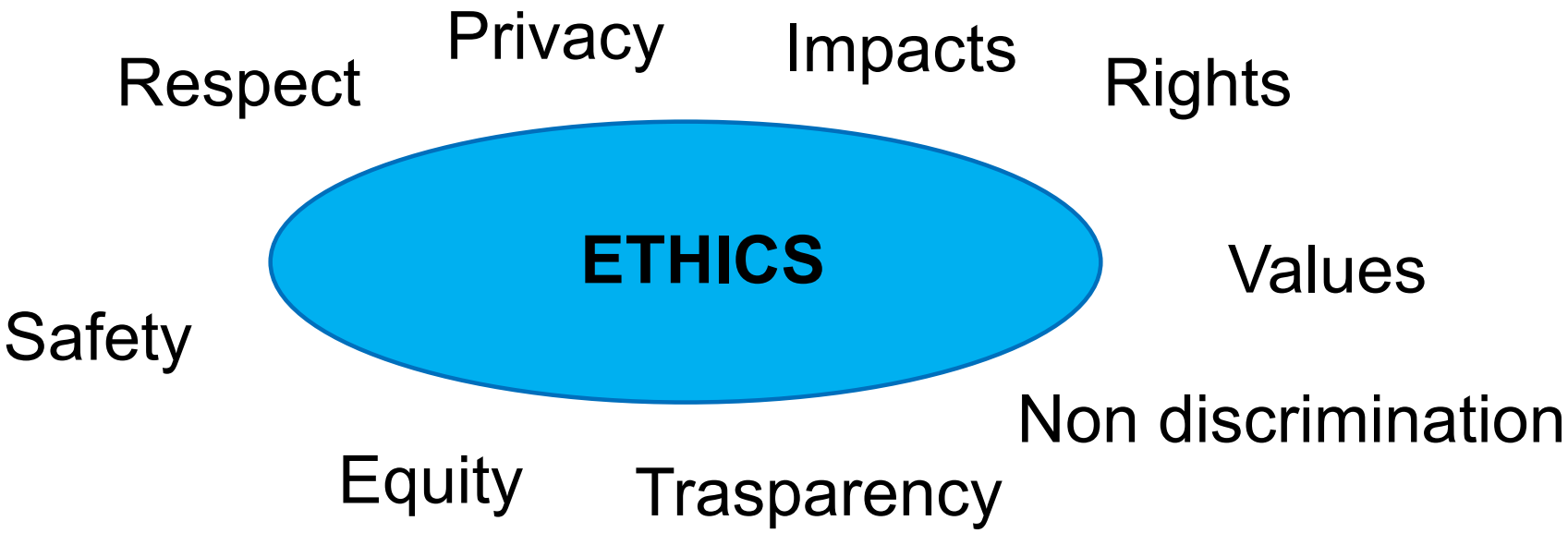
Evacuation planning: prescriptive model

Human factor



Plans should model the human aspects

Ethical factor



Plans should model the ethical aspects

Are ethical values universal?
Are ethical principles easily and clearly defined?

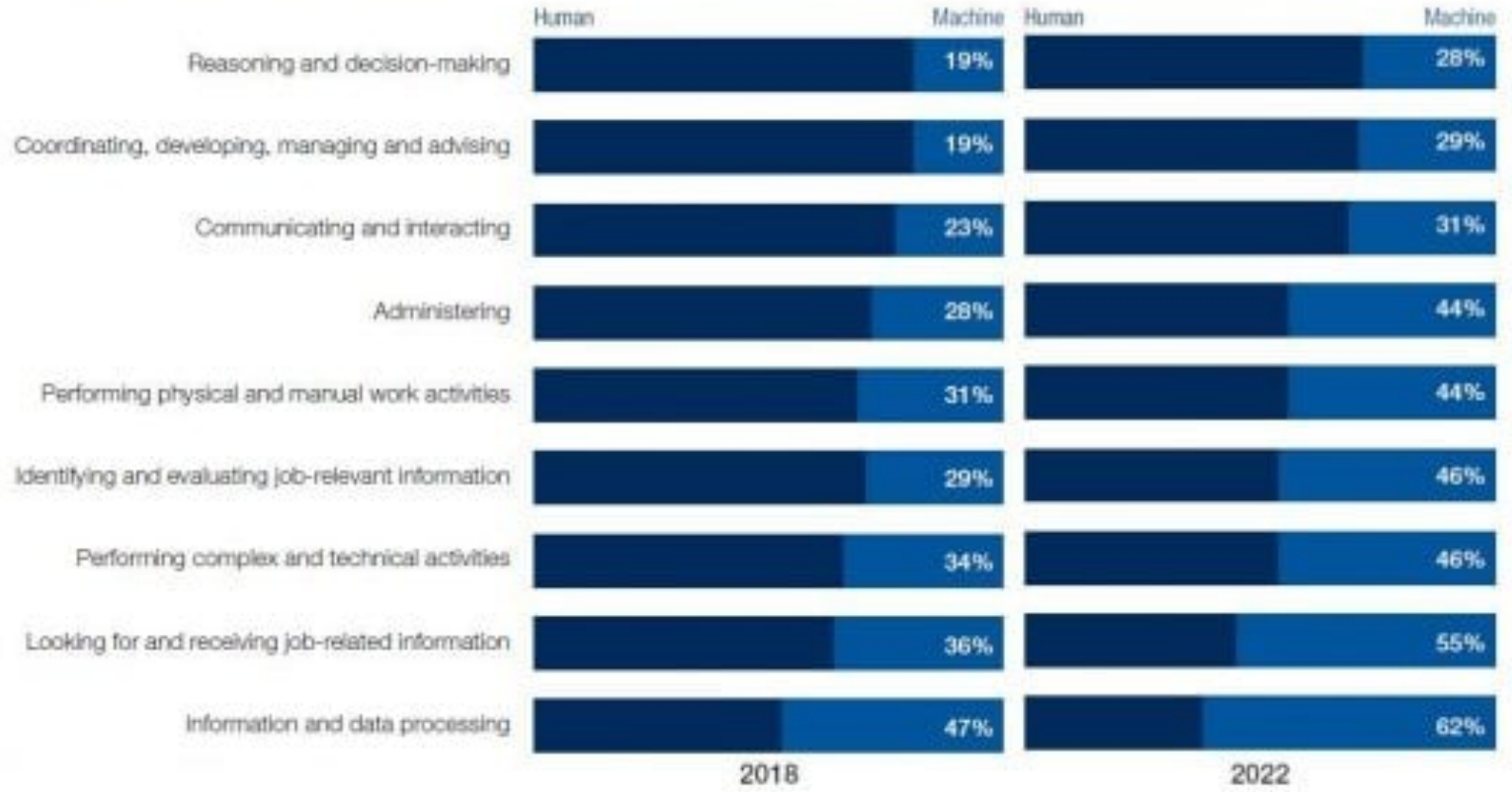
Application domains for AI

From “Artificial intelligence and life in 2030”, Stanford University – Sept 2016 – “AI eight domains with high impact:

- Trasports (intelligent cars, self driving cars, transport planning, on-demand transport)
- Domotics (companion robots)
- Health (clinic support, health data analisys, health robotics, elderly care)
- Education (tutoring system and on-line learning)
- Inclusion of poor classes
- Safety and security
- Job market
- Entertainment (social platforms, games, arts and creativity).

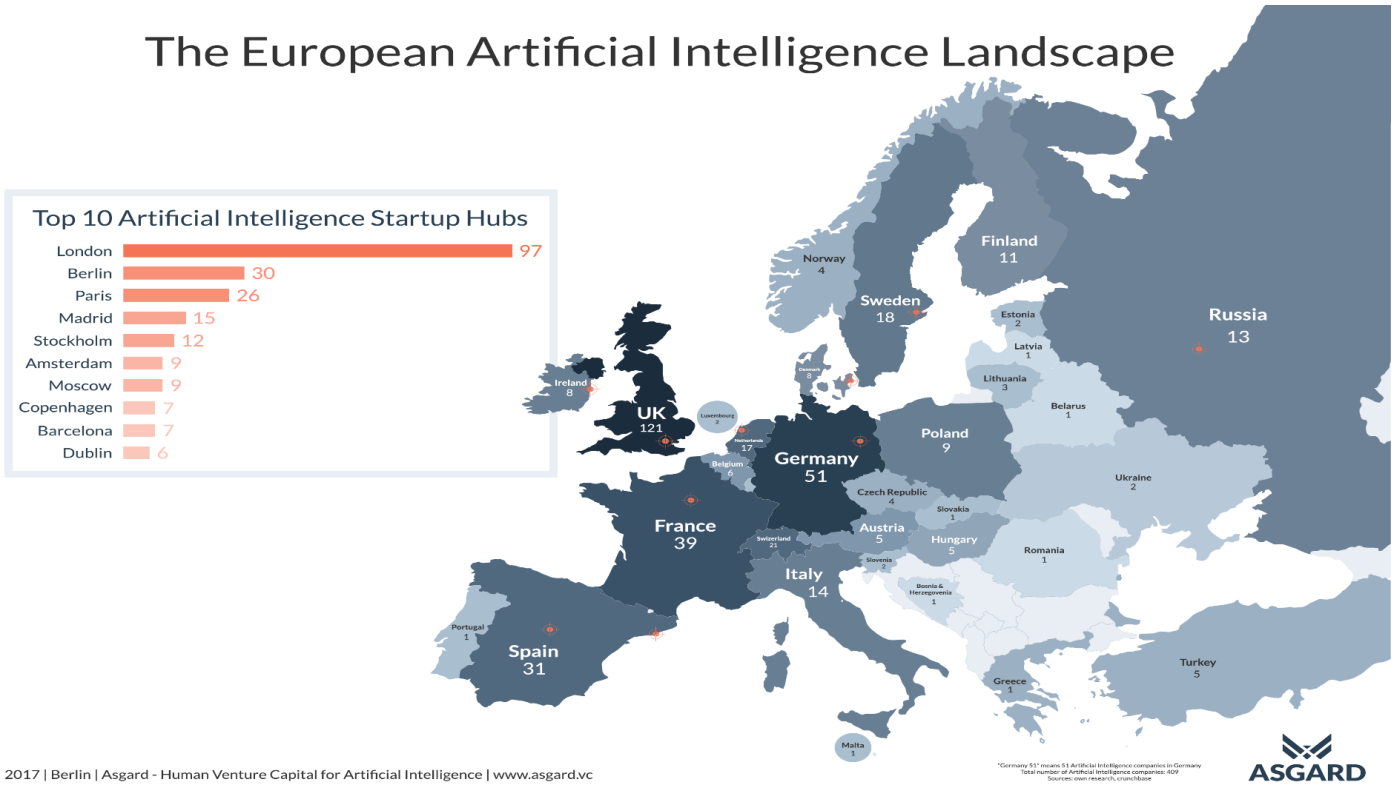
Ratio of human-machine working hours 2018 vs 2022

Figure 5: Ratio of human-machine working hours, 2018 vs. 2022 (projected)



Source: Future of Jobs Survey 2018, World Economic Forum.

EU landscape



Where AI

- European Association for Artificial Intelligence (EurAI), (previously ECCAI). Founded in 1982, is a scientific umbrella association organizing ECAI European Conference on Artificial Intelligence.

- Association for the Advancement of Artificial Intelligence (AAAI) (previously American Association for Artificial Intelligence). Founded in 1979, it is a scientific association organizing the AAAI Conference on Artificial Intelligence.

- Associazione Italiana per l'Intelligenza Artificiale (AI*IA). Founded in 1988, is a scientific italian association and organizing the Conferenza Italiana di IA.

- The largest AI conference in the world is the International Joint Conference on Artificial Intelligence (IJCAI). It will be organized in Bologna in 2022 together with ECAI. Voluntary needed !!!