Artificial Intelligence

→ Towards ‘Frankenmodels’

→ More efficient approaches towards AI systems

→ New ways to access data

→ Towards trustworthy AI

→ More accessible computing and models

Artificial Intelligence

Artificial Intelligence (AI) and, more specifically, machine learning (ML) have taken off in recent years. This resulted in impressive achievements, such as beating the world's best Go player with AlphaGO and near-lifelike image generation by systems such as Dall-E. However, it also led to worries and discussions about the possible risks and harm caused by AI systems.

AI technology is already impacting the research and educational sectors both in content and their operation and will undoubtedly continue to do so in the future. AI is a hot field of study among researchers, and machine learning is a valuable research method. Consequently, the range of applications is growing. Within education, AI is less mature in its implementation, as can be seen in the trend manifestations presented here. However, the promise of Artificial Intelligence Educational Devices (AIEDs) is growing both in the classroom (micro level), institution (meso level), and society at large (macro level).

AI is often deployed within a complex system containing the technological infrastructure within socio-technical contexts. This development is illustrated by trends covering technical themes such as computation, data, and hosting of services. It is also reflected in trends related to ways those technological elements should be deployed in specific communities to form a trustworthy ecosystem.

TREND #1

Towards 'Franken models'

Public values

Autonomy	Independence of education	
Justice	Equality | Integrity | Trustworthiness of information | Transparency | Democratic control | Sustainability | Inclusivity	
Humanity	Safety	

Readiness WATCH PLAN ACT

Drivers

#Dataism #Connectivity #Automation #Globalization #Open science

Early machine learning models were designed for a specific task. Recent developments show that we can now build more complex models that merge different kinds of input in a system with broader capacities. Innovations in architecture and multimodal, multi-objective training of models make this possible. Models are trained with different types of information and a diversity of tasks. Institutions with access to those models have a competitive advantage. These (Franken) models are often trained with weak labels or even unsupervised. The methodology involves using self-supervision, leading to generic capabilities. These models can be seen as a new pocketknife, unlocking new capacities and opportunities for automation, leading to value creation.

Foundation models

The term 'foundation model' has been coined by researchers at the Stanford Institute for Human-Centred Artificial Intelligence to capture the increase in large models trained at scale and adaptable to a wide range of tasks. These models all provide a platform of capacities that can later be tailored to specific applications such as BERT, GPT-3, chatGPT, DALL-E, and stable diffusion.

open example

Multimodal learning in healthcare

The multimodal approach helps researchers extract knowledge and value from an exascale dataset without requiring large labeled, annotated data. The supervision is done by using trained networks to provide labels that in turn are corrected through an active learning interface. We can combine images of tissues with their corresponding diagnostic text. A similar impact can be expected in fields where knowledge extraction relies on basic capacities from foundation models, such as combining data from sensors with multiple modalities or combining resources in a learning environment.

open example

DeepMind's Flamingo model

Flamingo is a visual language model (images + text) developed by DeepMind. It combines a visual and a language model, which means it can tackle a wide range of tasks related to those modalities. Also, Flamingo is a few-shot learning model, which means it can learn new tasks from just a few additional inputs. The next step is going from text to image to text to video applications. This technology has a big potential to help create engaging personal assistants for the education sector.

open example open example

IMPACT

These tools can unlock new capacities for both research and education: extract information from complex systems, generate new complex data, add new data modalities to existing pipelines, automate/accelerate existing tasks, and design more interactive and engaging courses. These models feature end-to-end capabilities, being able to model complex problems. As is often the case, those models are susceptible to bias, prejudice, and copyright issues. Non-uniform adoption of those tools may also increase efficiency gaps between relevant actors. Controlling the trained models will become more important to secure integrity. There is a potential risk for models to become too powerful and threaten our human values. We expect to see such generic, large, pre-trained models to be the starting point for future task design and training.

TREND #2

More efficient approaches towards AI systems

Public values

Autonomy	Independence of education
Justice	Integrity | Sustainability | Effectiveness
Humanity	

Readiness WATCH PLAN ACT

Drivers

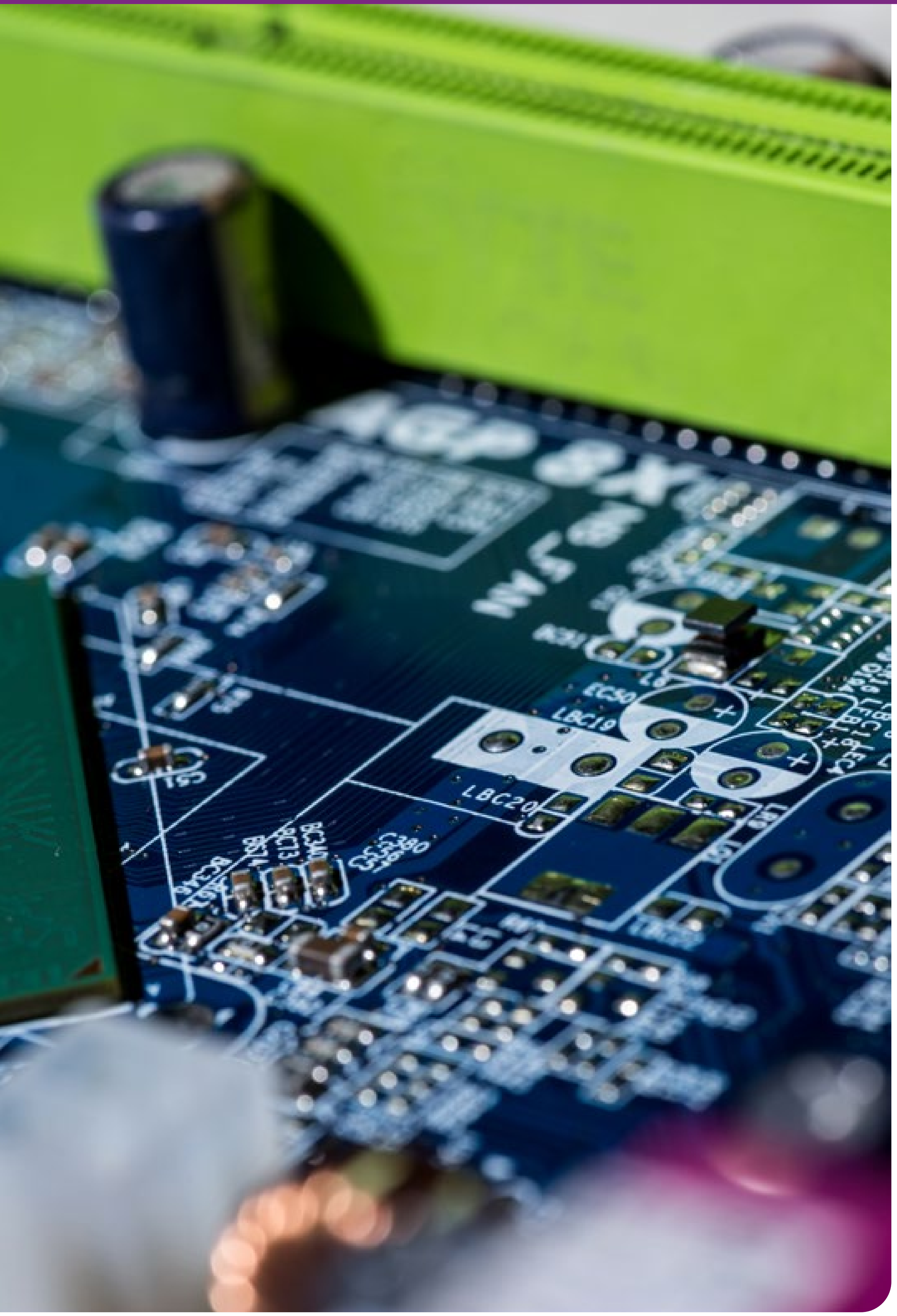
#Internationalization #Dataism
#Decentralization #Power efficiency
#Open Science #Open source development

As the use of AI systems increases, we see a trend toward ensuring the efficient training and deployment of AI systems. This trend is an effort to reduce both the data and computational cost of training and deployment of AI systems. Due to economic and environmental costs, it is not always feasible to train ever larger models. On the model level, inductive bias uses a priori knowledge of the studied system to constrain model architecture. A notable class of emerging networks is the spiking neural network, that promises to be more neuron efficient than a classical artificial neural network, i.e., achieving more complex behaviour using fewer neurons than traditional models. This is done by using biologically inspired activations. In addition to reducing the data and modeling complexity, there is also a move towards AI-specific hardware to effectively train and run models. Transitioning from CPU to GPU for certain computing tasks has led to even more specific hardware being developed.

Physics-informed neural networks

Multiple projects take physical laws into account (a form of inductive bias) when designing the simulation workflow to limit the data needs and model complexity. By doing this, the models automatically discard all possible ‘non-physical’ solutions to the problem, drastically reducing the number of outcomes and thus training time.

open example open example



Graph neural networks

GNNs are another example of inductive bias. These networks are particularly effective for tasks where data comes from non-Euclidian space (i.e., where the physical space is irrelevant). The most obvious example is of course social networks, but these networks can be applied in a surprisingly broad range of domains (e.g., natural language processing, chemistry).

open example

Towards AI-specific hardware

With the rise of AI there have also been new opportunities for alternative computational and data architectures to take advantage of the specific operations required by deep learning algorithms. From the early TPU from Google, many companies are now developing innovative solutions to train and deploy algorithms efficiently (e.g., Graphcore, Cerebras, Habana, Rain Neuromorphics).

open example open example

IMPACT

More effective AI systems will increase AI adoption by making them more accessible to a wider range of users. Inductive bias is likely to significantly impact research into AI and with AI. Moreover, the development of AI-specific hardware will give yet another boost to AI system capabilities and these capabilities will create demand for new infrastructures. The key driver here is to get the most out of limited resources. However, this requires highly specialised hardware and skills.

TREND #3

New ways to access data

Centralising and sharing data is not always feasible due to organisational complexity, excessive cost, or privacy concerns. Increasingly solutions are developed to gain access to insights from data that are difficult to share. This is done either through distributed training (federated learning), training on synthetic data, or secure data-sharing and computational environments. GDPR and other initiatives make data privacy top of mind, reinforcing this trend. The high (data) cost of training new models from scratch also is an incentive to look for creative ways to work with less data. This trend helps organisations retain control and ownership of data and may help to decrease the ecological impact of the AI ecosystem.

Federated learning

In fields where some data, or part of the data cannot be shared (e.g., health and education), it becomes crucial to have ways to train algorithms without having full access to the data. This is what federated learning solves by orchestrating a network of nodes, where each node partially learns an algorithm using only its data before aggregating the model centrally, without ever having access to the entire dataset itself.

open example

Public values

Autonomy	Independence of education
Justice	Integrity, Equal opportunity | effectiveness
Humanity	

Readiness WATCH PLAN ACT

Drivers

#Privacy #EU legislation #Digital economy #Data economy #Data governance #Automation

Synthetic data

Another approach to this issue is to generate a set of synthetic data based on a model of the true distribution and share that dataset instead. While powerful, this works only if a powerful base model to generate synthetic data can be trained in the first place. If a synthetic dataset can be used, this allows downstream applications to work on an equivalent dataset but without privacy issues. With state-of-the-art algorithms, such as the foundation models presented in the first trend, we expect a rapid development of synthetic data use in research and education.

open example

ODISSEI Secure Supercomputer

Through a secure supercomputing environment highly privacy-sensitive data of Statistics Netherlands (CBS) is made accessible for researchers in a safe and secure manner. This democratises access to population data and accelerates research within the social sciences.

open example open example

IMPACT

This trend is particularly helpful to provide AI capabilities to fields that until now have not been able to make full use of AI because of the difficulty to access data (e.g., health, education). By making it easier to share and access data, new research insights can be gained which will lead to new capabilities. However, the privacy risks linked to the synthetic data approach should not be underestimated. Once the appropriate infrastructure and protocols have been developed, synthetic data can both help maintain privacy and allow models to be trained on larger datasets. This will also create further possibilities for international research collaborations. Synthetic data also gives more control to data owners.

TREND #4

Towards trustworthy AI

Public values

Autonomy	Independence of education | Safeguarding of private life and personal data
Justice	Equality | Transparency | Democratic control
Humanity	Respect | Safety

Readiness WATCH PLAN ACT

Drivers

#EU legislation #Responsible technologies #Privacy #Carbon footprint #Globalization #Digital economy #Data governance #Diversity #Inclusion #Equity #Digital literacy #Gender balance

AI models' increased uptake and performance have led to more significant societal questions of responsibility, trust, and transparency. This is evidenced by the European Commission's focus on Trustworthy AI and upcoming AI legislation. Within trustworthy AI, several societal and technological developments come together to address these societal questions. Frameworks and legislation provide a baseline for the requirements AI should meet. A stronger focus on reliability, standards, and interoperability in development leads to more robust and transparent AI systems. Democratisation of AI leads to more resources to learn and more standards to follow, which also helps newcomers get started.

AI ethics guidelines

Within a growing discourse on the responsible use of AI, academics, NGOs and companies are discussing the impact of AI and its ethical implications. These discussions include topics like challenges of trust, transparency, fairness, and accountability. Many resources have emerged, but one of the most important is the ethical guidelines issued by the European Commission's High-Level Expert Group on AI.

open example

Algorithm registers

To provide transparency in the use of algorithms, multiple municipalities and governmental institutions have started using algorithm registers to provide the public insight into their use. In the future this is likely to become mandatory for all governmental organizations. Publicly sharing where, when, and how algorithms are used helps being transparent and accountable to relevant stakeholders such as citizens, users of applications, the media, and the relevant authorities.

open example

Standardisation and professionalisation

From data acquisition and curation to training and deployment, the typical engineering workflow is complex. As with software engineering best practices and tools, there is a trend in AI led by major companies and research groups to facilitate such workflows through, for example, MLOps. Standardization of operations and deployment leads to more mature, professional environments. This is crucial for interoperability but also helps with transparency to promote trust and avoid biases.

open example

Community-led initiatives

Large AI models are often only accessible to large research labs and corporations. Community-led efforts such as BLOOM aim to democratise AI models by making them available for researchers in smaller labs with fewer resources. These efforts also involve more parties in the training phase to insure better transparency and more robust models. BLOOM is a collaboration of 1000+ researchers from 70+ countries, which trained a large language model for 46 natural languages and 13 programming languages.

open example

IMPACT

Trustworthy AI is a critical driver that impacts all parties working on AI, either in research or education. Formalising good practices and standards helps professionalize the community and provides resources for responsible use and fast, practical tools. Sufficient attention to trustworthy AI can prevent that (1) damage is done to communities through irresponsible use of AI, and (2) communities are split with different standards on how to work with AI methods.

TREND #5

More accessible computing and models

AI methods are becoming a scientific instrument that can be used as a readily available tool, the same way as a microscope. As a 'new tool' among others, AI must be accessible as a commodity at a low cost to provide easy access to non-experts. As most deployments take place in cloud environments, AI is increasingly accessible in the public domain. The use of abstraction layers (such as autoML or MLaaS approaches) helps to provide direct access to complex computational abilities for non-technical expert stakeholders. However, this leads to a trade-off between ease of access and the risk of vendor lock-in and privacy concerns when using these applications in commercial cloud infrastructure.

Public values

Autonomy	Freedom of research and education
Justice	Equality | Inclusivity
Humanity	

Readiness WATCH PLAN ACT

Drivers

#Research environment #Digital literacy #Dataism #Decentralization #Automation #Connectivity #Digital economy #Globalization

Low-code/no-code

With more advanced software layers on top of computing infrastructure, there is a low/no code movement. Several service applications enable complex computational tasks with little or no programming. While traditional HPC centres still have an edge for advanced intensive research, low/no code makes computing more accessible to new groups of stakeholders.

open example

Toward machine learning as a service (MLaaS)

Related to the previous example, the growth of platforms to share datasets, pre-trained models and easily build on them is a good example of the ongoing democratization of AI methods and the lower entry costs. These platforms also help with transparency and standardization of methods.

open example

Facilitating model design with autoML

An important part of the current data scientist workflow is to experiment with model architecture for a given problem and to tune the training procedure to obtain the best results. This is very time-consuming and highly dependent on the practitioner's experience. Instead, autoML proposes to automate this part of the workflow and leave more added-value tasks to the user. Examples are how to frame the problem of interest, what metric is relevant and how to interpret results in a responsible way.

open example

IMPACT

This trend is a double-edged sword: it democratizes access to computing resources and AI methods but at the expense of potential vendor lock-in, privacy issues, and irresponsible use of those algorithms due to a lack of in-depth knowledge. Research and educational institutions will need to learn and follow guidelines for the responsible use of AI and understand the implicit trade-offs when choosing different computing platforms. The democratization of AI methods should help progress in fields that were left out until now.

More about Artificial Intelligence

Contact

Matthieu Laneuville
Program manager AI Lab
matthieu.laneuville@surf.nl

Damian Podareanu
Team lead High-Performance Machine Learning
damian.podareanu@surf.nl

Duuk Baten
Advisor Responsible AI
duuk.baten@surf.nl

Bertine van Deyzen
Project manager AI in Education
bertine.vandeyzen@surf.nl

Continue reading

more info

communities