Steering human-centred use of generative AI in education

Fengchun Miao

Chief, Unit for Technology and AI in Education
Future of Learning and Innovation, Education Sector, UNESCO

Human-centred approach to AI

AI supporting human capacities;
AI are explainable, predictable, human-controlled, human-accountable,
capable of being shut down

Inclusion and equity

Humanistic principles

Human rights and human dignity

Supporting the formation of national digital learning policies (80+ countries)

5 UNESCO policy guidelines for digital learning

AI and Education: Guidance for Policy makers

Guidelines for ICT in Education Policies and Masterplans

Guidance for generative AI in education and research

Guidelines on the Development of OER Policies

Education and Blockchain

Continued support for digital learning policies

Worldwide grass-rooted case studies

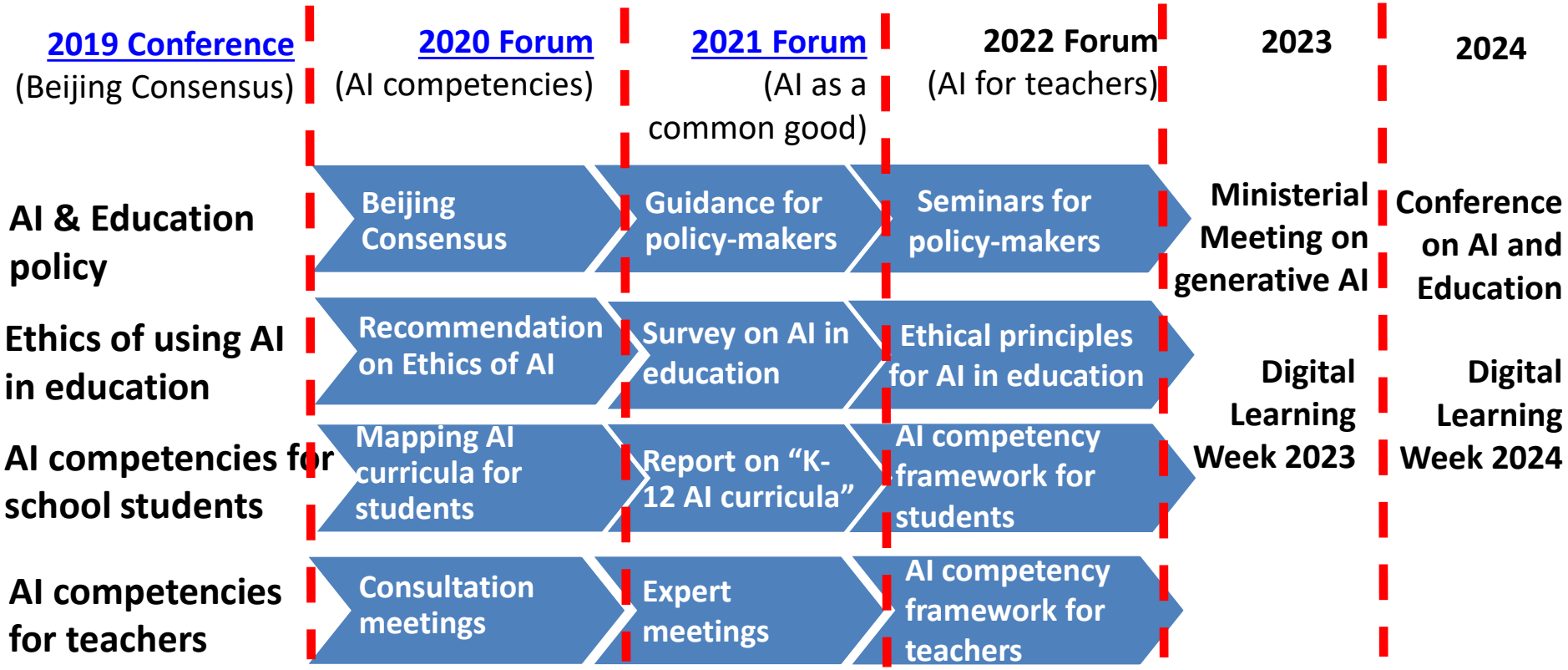
- 80+ examples on AI & ED: Compendium (2019) ; Compendium (2020)
- First global mapping of government-endorsed AI curricula (2022)

- 24 winning projects of UNESCO Prize for ICT in Education
- Case studies on the distance learning in response to COVID-19 disruptions: Finland, Korea, Saudi Arabia
- 14 best practices in mobile learning

- OER: Policy, costs, and transformation **–14 frontrunner projects on OER**

UNESCO Work on AI and Education

(100+ ministers and 10 000+ participants from 120+ countries)

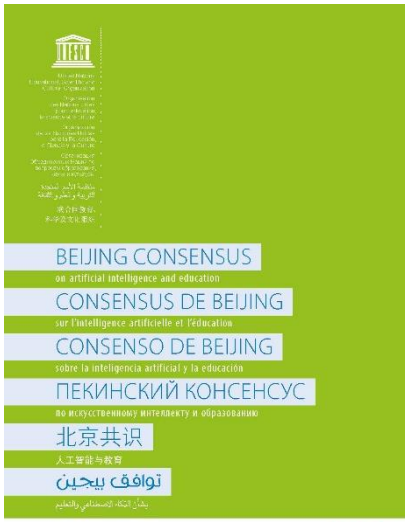


Guidance for generative AI in education and research

A publication on "AI and Futures of Learning"

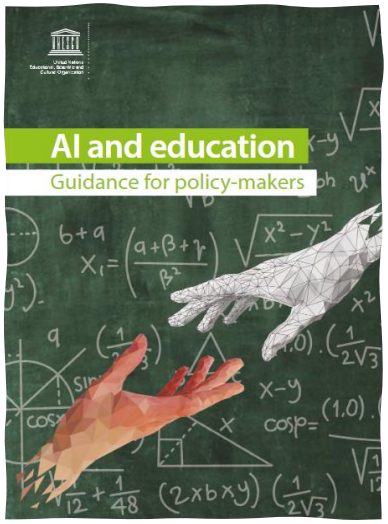
Main publications

Top 20 for 1 year



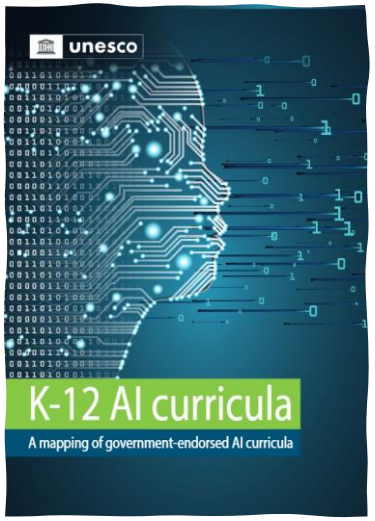
6 UN languages

Top 10 for 1 year



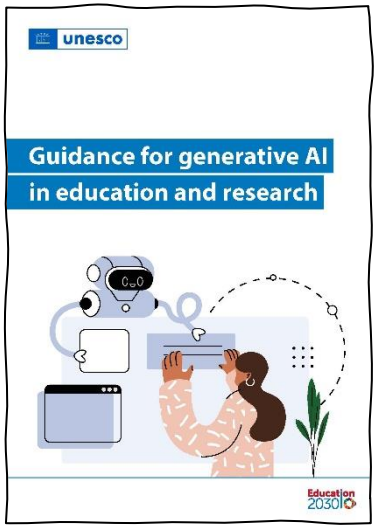
6 UN languages + Korean

Top 40 for 1 year



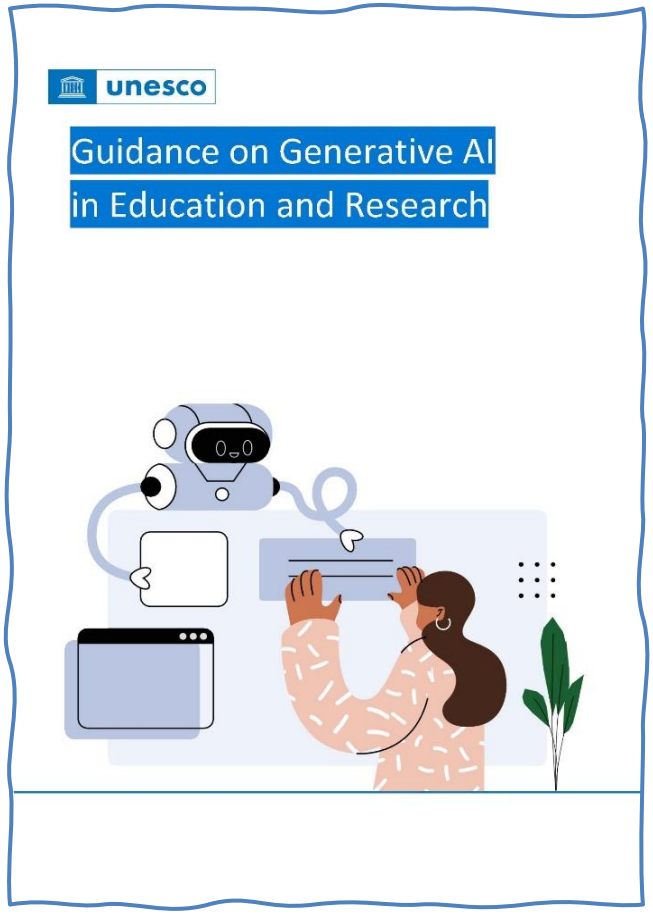
6 UN languages + Portuguese

No.1 for 2 months



Will be available in 10+ languages

Guidance for generative AI in education and research



Contents of the Guidance

- 8 controversies
- 6 steps to regulate generative AI
- Policies on building capacities
- Institutional validation
- Design of uses
- Reflection on long-term implications

1. GenAI: definition and potentials

1.1 Definition

- An AI technology that automatically generates new content **across all symbolic representations of human thinking**: natural language, images (photos, digital paintings and cartoons), videos, music and software code.
- It generates its content by **statistically analysing the distributions of words, pixels or other elements in the data** that it has ingested and identifying and repeating common patterns

1.2 Tech leaps

Combining multiple deep learning models, the optimization of architectures, the refining of parameters at billions scale, crawling of data across platforms, exponential growth of computing capacities → foundation models→ transformation at digital infrastructure level.

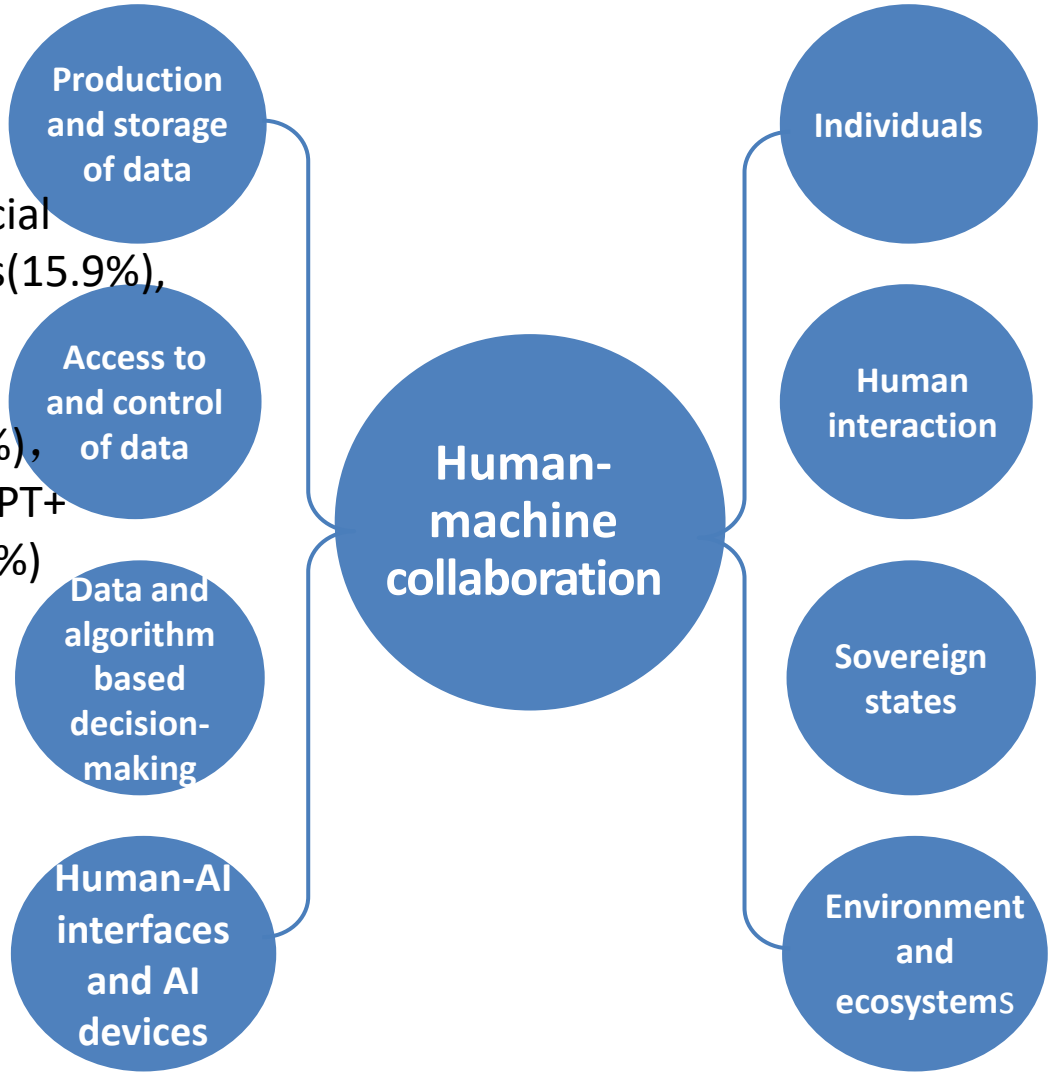
1.3 GenAI for education: Three assumptions

- Trilemma of AI innovation, trustable AI, and inclusive use of AI → a tri-symmetric advancement
- GenAI directly threatening core values of education (equity, inclusion, human agency in learning, values and linguistic diversity, pluralism) → examine GenAI against core values
- Prioritizing regulations → Ensuring inclusion → Building capacities for proper uses

1.4 A human-AI collaboration model for examining GenAI

AI: from moving targets to proportionality

- Data sources: Crawling webpages (61.75%), Social media (8.86%); Libraries(15.9%), wikipedia (3.49%)
- Languages: EN (92.647%), FR(1.819%), DE+ES +IT+PT+ NL (2.60%), …CN (0.099%)
- Unexplainable methods to generate outputs
- Generative AI doesn't understand semantics and the real world



Education: knowledge evolution

1. Evolution of knowledge production
2. Projection social values
3. AI job skills; challenging learning outcomes

2. Eight controversies around GenAI

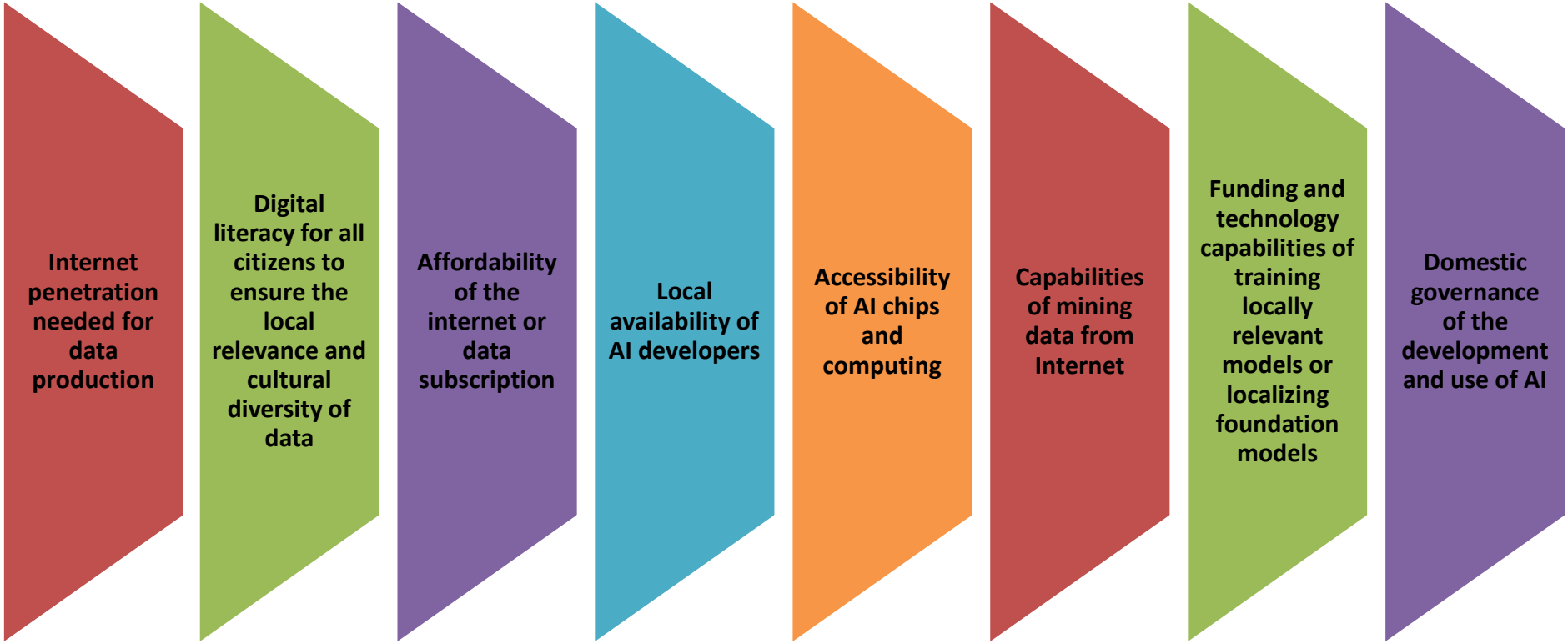
Q1

- Do we allow data of the poor to be deprived without pay and the poor people need to pay for AI services generated on their data?

Data poverty and inclusion

Controversy 1:
Worsening digital poverty

A taxonomy of data poverty



→ *A tax on the mining of data to train GenAI*

Q2

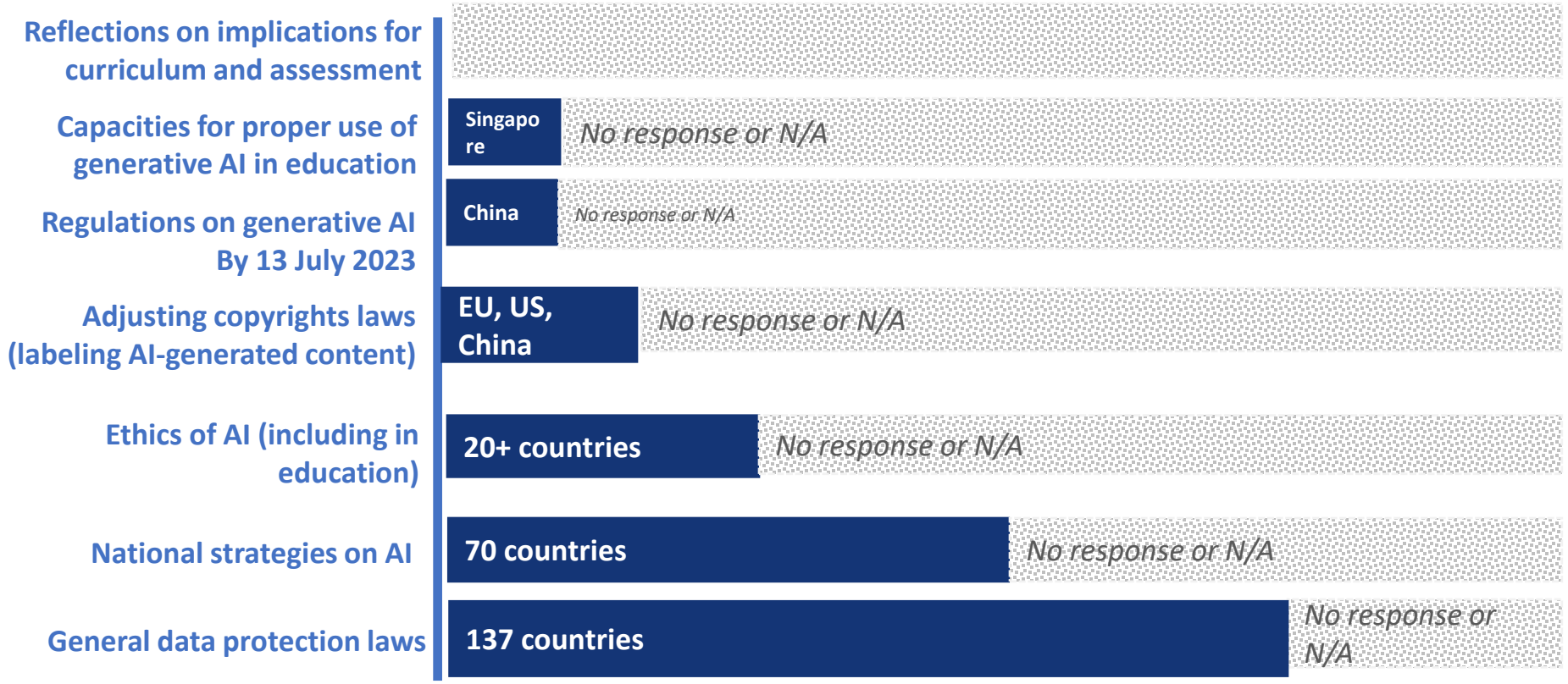
- Do we accept AI iterations outpacing national regulation adaptation?

Undermining domestic control

→ Regulations to regain domestic control

Controversy 2:
Outpacing national regulatory adaptation

Tiered shortage of regulations on AI and GeAI in particular



Q3

- Do we accept generative AI using copyrighted content without consent?

Data ownership and copyrights

→ adjust copyright laws on use of content

Controversy 3:
Use of content without consent

- Violating GDPR
- New York Times lawsuit against OpenAI
- Violating users' **right to be forgotten** (GenAI models can't unlearn the identified patterns)

Q4

- Do we believe finding answers by clicks and prompts from unexplainable models is learning and can help intellectual development?

Unexplainable models

for learning, thinking, and intellectual development

Controversy 4: Unexplainable models used to generate outputs

The Foundation Model Transparency Index (Stanford University)

Indicators / Examples of foundation models	Llama2 （%）	ChatGPT-4 （%）	PaLM 2 （%）
Data	40	20	20
Data labor	29	14	0
Compute	57	14	14
Methods for model's training	75	50	75
Model basics	100	50	67
Model access	100	67	33
Capabilities	60	10	80
Risks	57	57	29
Model mitigations	60	60	40
Distribution	71	57	71
Usage policies	40	80	60
feedback	33	33	33
Impact	14	14%	0
Average scores	57	47	39

Q5

- Do we accept children chatting in platforms trained mainly for adult users?

Age restrictions

→ for independent conversations with GenAI platforms

Controversy 5:
AI-generated content polluting the internet

- Disinformation, misinformation, hate speeches
- Continuous crawling online data polluted by AI generated content (identified patterns) models will lower the quality of data for training models and can lead to collapse of models

Q6

- Is GenAI designed to provide trustable learning content and reliable formative tutors?

Can identify syntax, but lack understanding

of sementics and the real world

Controversy 6:
Lack of understanding of the real world

GitHub's Hallucination Leaderboard

Model	Accuracy	Hallucination Rate
GPT 4	97.0 %	3.0 %
GPT 4 Turbo	97.0 %	3.0 %
GPT 3.5 Turbo	96.5 %	3.5 %
Llama 2 70B	94.9 %	5.1 %
Llama 2 7B	94.4 %	5.6 %
Llama 2 13B	94.1 %	5.9 %
Anthropic Claude 2	91.5 %	8.5 %
Google Palm 2	87.9 %	12.1 %
Google Palm 2 Chat	72.8 %	27.2 %

Q7

- As generative AI projects data holders' cultural views, shall users from marginalized groups forego cultural diversity when accepting AI services?

Constraining plural opinions and cultural/linguistic diversity

→ Encourage plural opinions and expressions

Controversy 7:
Threatening plural knowledge construction

- Data-poor populations have limited digital presence online → Their voices are not represented in the data → GPT models can further marginalize already disadvantaged
- LLMs are rife with different political biases
- Bard, ChatGPT, GPT-4, Claude all present perpetuating race-based values in their responses to neutral Qs in medicine.
- ChatGPT replicates gender bias in recommendation letters
- The diversity of ideas among participants who used GPT-4 for the creative product innovation task was 41% lower compared with the group that did not use the technology.

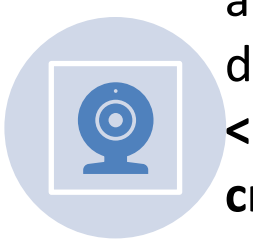
Q8

- As generative AI outputs can subvert human and technology detection, will be fake papers, news, and images accepted?

Deeper deepfakes

→ Copyright laws on Deep Synthesis

Controversy 8: Generating deeper deepfakes



Nearly zero-cost generation and sharing of deeper deepfakes especially: **taking <25 mins and nearly $0 to create a 60-second deepfake pornographic video** using just one clear face image.

The total deepfake videos by October 2023 presents a 550% increase over 2019 with 98% being deepfake pornography and 99% of targeting are women.

https://www.homesecurityheroes.com/state-of-deepfakes/#key-findings

3. Proportionality of GenAI for education

3.1 Safety for non-GenAI supplier countries

- Digital safety under threat
- GenAI doesn’t understand values, but project values
- Irrelevance for local languages and curriculum

Ensure safety of GenAI

- Enhance domestic governance
- Autonomic foundation models for education

3.2 Age appropriateness of chat-based learning for children

Age limitation ***for independent conversations with GenAI platforms (13 years or 16 years)***

3.3 Pre-cooked content for learning

3.3.1 Content hallucinations lead to knowledge illusions

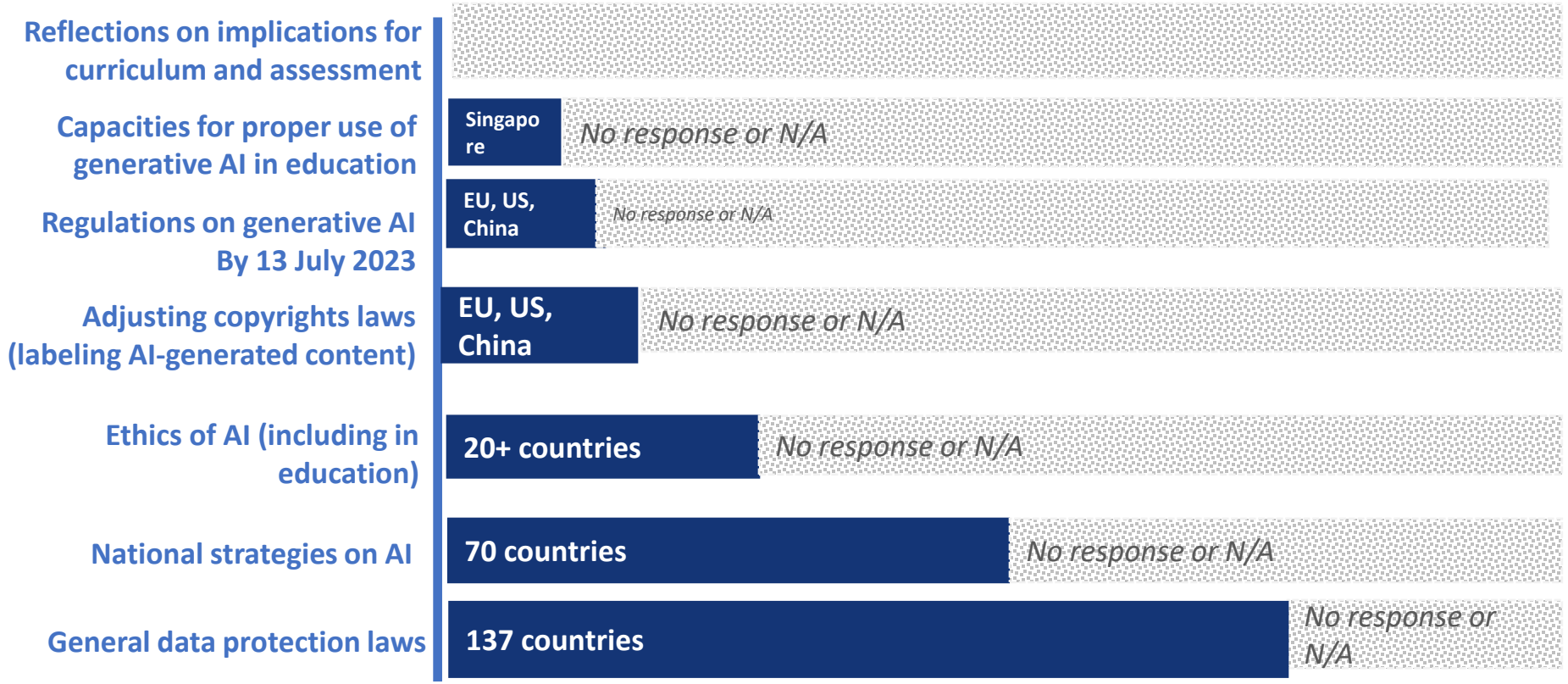
3.3.2 Outsourced writing replaces learners' training of thinking and intellectual development

3.3.3 Pre-cooked processing of information lead to inert innovation and intelligence enfeeblement

4. Strategies

4.1 Regulate AI (and GenAI in particular)

- Risk categorization
- International regulations
- Intersectoral co-governance



4.2 Validate AI for education and protect human agency

- Promote inclusion, equity, linguistic and cultural diversity
- Protect human agency
- Ethical and pedagogical validation
- AI competencies for learners
- AI competencies for teachers and researchers
- Promote plural opinions and expressions
- Local foundation models AI application
- Review long-term implications

4.3 Build AI competencies for teachers and students

UNESCO AI competency framework for **teachers**
(under development)

Aspectsc	Progression		
	Understand	Apply	Create
Human-centred Mindset	Understanding benefits & risks	Contextual strategies	Steering long-term impact
Ethics of AI	Human agency	Human-centred use	AI society skills
Foundation AI knowledge	“Algorithm and data literacy” or AI literacy	Use AI analytics	Coding and data models
AI skills	Test and use	Infusing uses	Integrating AI tools
AI pedagogy	AI for teaching	AI to deepen learning	AI for co-creation
Professional development	AI to assist administrative tasks	AI for curriculum design and delivery	AI empowering teaches

4.3 Build AI competencies for teachers and students

UNESCO AI competency framework for **stuendents**
(under development)

Aspects	Progression		
	Understand	**Apply**	**Create**
Human-centred mindset	Critical views of AI	Safe use	Self-fulfillment in AI era
Ethics of AI	Understanding human agency	Responsible use	AI society skills
Foundational AI knowledge	Algorithm literacy Data literacy	Coding and data processing	Algorithmic data analytics
Application skills	Domains and technique of AI	Practical skills	Modifying or creating tools
***Design thinking for problem solving**	Problems abstraction	Model architecture	Co-creation

Upholding core humanistic values when using GenAI

- human agency
- inclusion
- equity
- gender equality
- linguistic and cultural diversities
- plural opinions and expressions

4.4 Guide instructional design on uses of GenAI

A “Human-agent and pedagogically proper interactions” template

Potential but unproven uses	Appropriate domains of knowledge or problems	Expected outcomes	Appropriate GenAI tools and comparative advantages	Requirements for the users	Required human pedagogical methodologies and example prompts	Possible risks
AI Advisors for augmented research outlines						

Co-design uses of GenAI in education

Human potentials and human touch	*Human-centred uses of AI*
Research	▪ AI advisors for augmented research outlines ▪ Generative data explorer and literature reviewer
Teachers and teaching	▪ Curriculum, lessons, and assessment co-designer ▪ Generative twins of teachers' assistants
Self-paced acquisition of basic skills in languages and arts	▪ 1:1 art coach ▪ 1:1 language skills coach ▪ 1:1 writing coach
Inquiry or project-based learning	▪ Socratic collaborator ▪ Facilitator for project-based learning
Support learners with special needs	▪ Inclusive learning ▪ Conversational diagnosis of learning difficulties

Potential use models of GenAI for teaching and learning

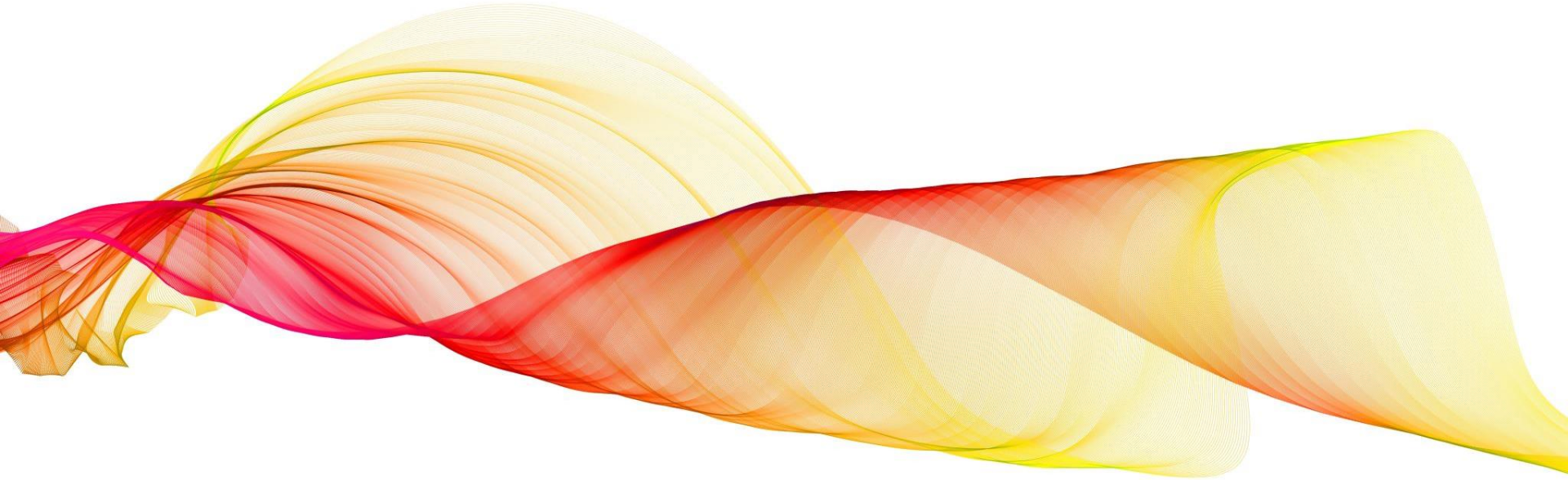
- **Foundation models** as digital infrastructure for learning →***Local EdGPT***
- **Transformers** for curriculum co-design across text, images, sounds, videos → ***Curriculum co-designers***
- **Formative discriminators** for computational performance of basic skills learning → ***AI twins of teacher assistants***
- **Superchargers** for improving the efficiency of higher-order thinking or inquiry-based learning →***AI assistants for research and inquiry-based learnings***

A fundamental ask

- As generative AI can automate the creation of essays and artworks, what should students be asked to learn?

Long-term implications

→ Rethinking learning outcomes



Long-term implications

- **Uncharted ethical issues**
- **Copyright**
- **Sources of content and learning**
- **Homogenized responses versus plural outputs**
- **Rethinking assessment and learning outcomes**
- **Thinking processes and intellectual development**

Rethinking learning outcomes

- **Values** to ensure the human-centred design and use of AI.
- **Foundational knowledge and skills** adapted to the increasingly AI-rich environments.
- **Higher-order thinking skills** based on human-AI collaboration and generative AI outputs.
- **Vocational skills needed to work for and with generative AI**.

Thanks!

f.miao@unesco.org

https://www.unesco.org/en/digital-education

https://www.linkedin.com/in/fengchun-miao-5b999077