Symbolic Computation, Standardisation, and the EU AI Act

James Davenport

University of Bath
ISO-IEC JTC1 SC42 JWG5, WG3 etc.
IEEE P.3109 Study Group
CEN-CENELEC JTC21 WG3 Convenor
Partially supported by STANDICT
All views expressed are personal

24 September 2025

REGULATION (EU) 2024/1689
OF THE EUROPEAN PARLIAMENT
AND OF THE COUNCIL
of 13 June 2024
(but drafting preceded the ChatGPT era)

- 44 pages of 180 recitals
- 74 pages of 101 articles
- 5 pages of supplementary articles
- 20 pages of Annexes

Generally enters into force 2 August 2026 (but prohibited uses on 2 February 2025, and some other special cases).
Written as "product safety" legislation.
Many more systems/uses of systems fall/may fall into the "high risk" category (Annexe III) than one might believe.

Key Definitions

provider means a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge;

deployer means a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity;

These pervade the Act, but aren't in general use outside this framework. A provider might also be a deployer.

~~SC, Standardisation, and t~~he EU AI Act: Article 13 Transparency and provision of information to deployers

1. High-risk AI systems shall be designed and developed in such a way as to ensure that their operation is sufficiently transparent to enable deployers to interpret a system's output and use it appropriately. An appropriate type and degree of transparency shall be ensured . . .
2. High-risk AI systems shall be accompanied by instructions for use in an appropriate digital format or otherwise that include concise, complete, correct and clear information that is relevant, accessible and comprehensible to deployers.
3. The instructions for use shall contain at least the following:
   (b) the characteristics, capabilities and limitations of performance of the high-risk AI system, including:
   (ii) the level of **accuracy**, including its metrics, **robustness** and cybersecurity[. . .]against which the high-risk AI system has been tested and validated and which can be expected, and any known and foreseeable circumstances that may have an impact on that expected level of accuracy, robustness and cybersecurity;

~~SC, Standardisation, and t~~he EU AI Act: "accuracy"

JHD Consider a cancer which occurs in 1%. Two tests exist.

A. Always no cancer.
B. "Possible cancer" if there, but also on 2% of cases where not present.

A is more accurate (1% error rate, versus B's 1.98% error rate) but completely useless.

EC (§2.6 of Annex to standardisation request)

"accuracy" shall be understood as referring to the capability of the AI system to perform the task for which it has been designed. This should not be confused with the narrower definition of statistical accuracy, which is one of several possible metrics for evaluating the performance of AI systems.

"Standardisation" is a phrase that many use, but comparatively few understand. Most developed, and many other, countries have standardisation bodies (generally one, Germany has two).

USA	ANSI = American National Standards Institute.
UK	BSI = British Standards Institute
France	AFNOR = Agence Français pour la NORmalisation
Romania	ASRO
Germany	DIN and VDE. (BSI is cybersecurity agency)

In general, these are independent bodies, though they can receive "requests" from their national government.
These are members (full members for developed countries, but various "associate" status are possible) of international standardisation bodies.

International Standards Bodies (relevant)

IEC The International Electrotechnical Commission held its inaugural meeting on 26 June 1906

ISO The International Organization for Standardization was founded on 23 February 1947

Is computing an electrotechnical subject or not? Lengthy discussions, culminating in

JTC 1 Joint Technical Committee 1, entitled "Information technology", which was created in 1987

SC42 Sub-Committee 42, entitled "Artificial Intelligence", was created in 2017

JWG5 Joint Working Group ISO/IEC JTC1/SC42 - ISO/TC 37 WG: Natural language processing

European Standards Bodies (relevant)

CEN European Committee for Standardization (French: Comité Européen de Normalisation) was founded in 1961. 34 member countries (including UK)

CENELEC European Committee for Electrotechnical Standardization (French: Comité Européen de Normalisation ÉLECtrotechnique) was founded in 1973 (as a merger)

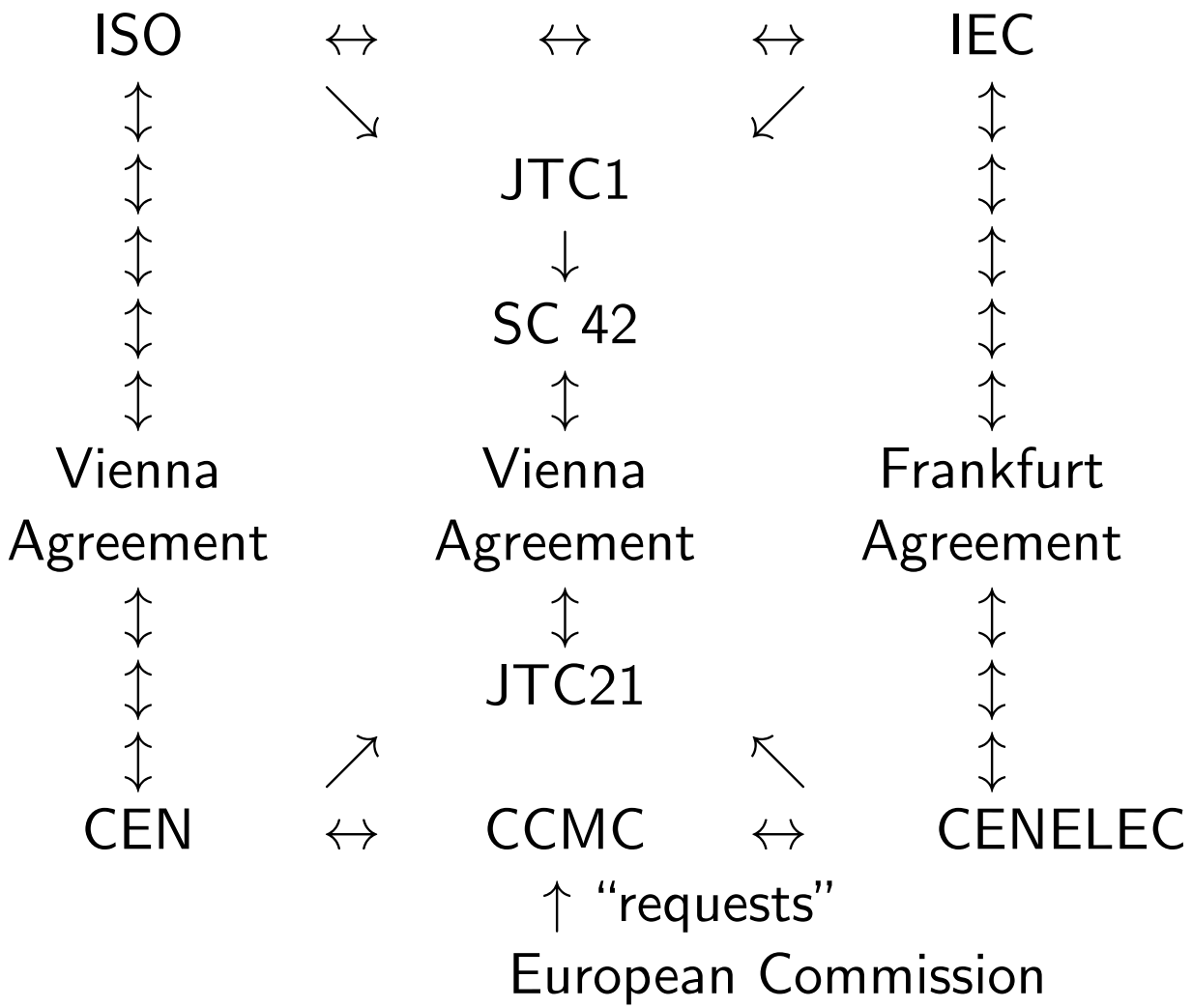
CCMC CEN–CENELEC Management Centre

ETSI European Telecommunications Standards Institute was set up in 1988 by the European Commission, and has a very different constitution.

Is computing an electrotechnical subject or not? Lengthy discussions, culminating in many Joint Technical Committees, including:

JTC 21 Joint Technical Committee 21, entitled "Artificial Intelligence", which was created by CEN & CENELEC in 2021.

Structure


ISO
IEC
JTC1
SC 42
Vienna Agreement
Vienna Agreement
Frankfurt Agreement
JTC21
CEN
CCMC
CENELEC
"requests"
European Commission

~~SC, Standardisation, and the EU AI Act:~~ Example

Car seat belts (a policy issue, but technical details): UK example

- 1955± Discussions in BSI about seat belts.
- 1960 BS 3254 Specification for Seat Belt Assemblies for Motor Vehicles.
- 1966 Compulsory in front seats of all new cars (certified with designated approval mark, i.e. BSI)
- 1983 Front seat wearing compulsory.
- 1987 Compulsory in all seats of all new cars
- 1991 Compulsory wearing in all seats; 3254:1991 replaces 3254:1988 which replaced . . .
- 2002 BSI 3254 replaced by UN/ECE Regulation 16

Standards Publications/ Levels

There are various types of international/european standards (and many national ones!)

TR Technical report. No requirements. Often definitions, descriptions of issues etc.

TR 22989:2023 Information technology — Artificial intelligence — Artificial intelligence concepts and terminology.

TS "A Technical Specification addresses work still under technical development, or where it is believed that there will be a future, but not immediate, possibility of agreement on an International Standard. " [ISO/IEC, not CEN]

IS International Standard. Can contain requirements: "shall" etc.

EN European Norm. Equivalent of IS. Many IS are "adopted" or "adapted" as ENs.

hEN "harmonised EN", published on OJEU with Annex ZA explaining which clauses of the standard give a presumption of conformity with which pieces of European legislation.

~~SC, Standardisation, and the~~ EU AI Act: 13.3(b)(ii) "The level of accuracy"

There will be a (multi-part) standard addressing Article 13, intended to be a hEN.
The "accuracy" clause will repeat §2.6: "accuracy" shall be understood as referring to the capability of the AI system to perform the task for which it has been designed.
So we need "tasks" as well as "metrics".

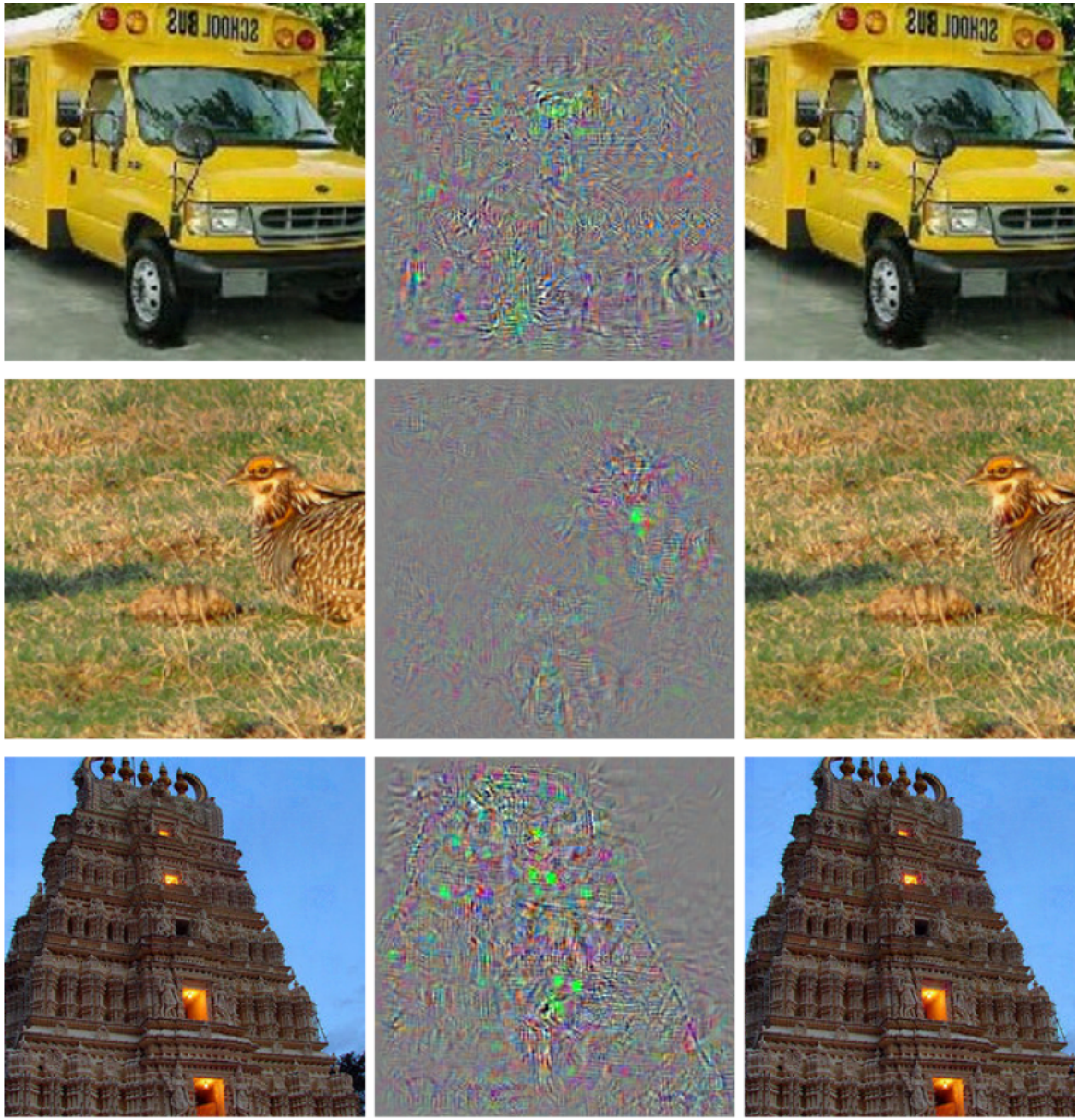
Many ISO-IEC 4213 — Performance measurement for AI classification, regression, clustering and recommendation tasks.

Vision Image recognition etc. Standards being developed in JTC21/WG3: a TR "taxonomy of tasks" and an EN for metrics of accuracy.

NLP Standards in JWG5: a TR 23281 "taxonomy of tasks" and an IS 23282 for metrics of accuracy.

Ideally, a standard has a nice simple requirement, and a corresponding method of testing it.
"A seat belt clip shall not open under a force of NNN Newtons, distributed in any way across the two straps. This will be tested by NNN Newtons on each strap, and by an equally distributed test".
But there is the "teaching to the test" problem, which is why human natural language examinations have "unseen translations" etc. See also "Dieselgate", where cars were, essentially, programmed to recognise the test track.

Robustness: adversarial images



Adversarial examples generated for AlexNet in [SZS+13].
Left image is normal, centre image is difference magnified by 10x, right is predicted to be an "*ostrich, Struthio camelus*" .
"large bounds do not automatically translate into existence of adversarial examples; however, small bounds guarantee that no such examples can appear."

Truth, Lies and Bullshit

Truth-teller	Knows the truth and tells it.
Liar	known the truth, but tells something different
Bullshitter	[Fra05] doesn't know/care about the truth, but says what is convenient.

I have a colleague who describes all generative tools as AB (Artificial Bullshit) rather than AI. This may be a better expression than "hallucinations".

What can we do about this: I personally distinguish "internal" methods from "external" ones. Internal robustness is considered by ISO-IEC 24029 series.

Gowers [Gow24] variant on wolf/goat/cabbage [Alc96]

Q A farmer needs to cross a river with two chickens. The boat has room for a human and two animals.

1. The farmer takes both chickens across
2. The farmer returns alone
3. The farmer takes one chicken across the river

ChatGPT Situation: Farmer and one chicken on the other side, one chicken on stanrting side.

4. The farmer returns alone
5. The farmer takes the second chicken across the river

ChatGPT So the smallest number of crossings needed is 5 crossings.

No underlying model!

SC, Standardisation, and the EU AI Act: 24029–1 Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview. [ISO21]

It was a mistake to call the standard "neural networks".

Definition (3.6)

robustness ability of an AI system to maintain its level of performance under any circumstances.

In practice, we distinguish

Robustness protection against what can go wrong:

* "Murphy's Law": what can go wrong, will go wrong.

Cybersecurity protection against what can be made to go wrong:

"Programming Satan's Computer" [AN95]

SC, Standardisation, and the EU AI Act: 24029–3 [ISO25] AI — Assessment of the robustness of neural networks — Part 3: Methodology for the use of statistical methods

It was a mistake to call the standard "neural networks" (and JTC21 will find a way of dealing with this).
Most of these methods are based on seeing what happens to measures of accuracy.
Since they are statistical, they will probably not catch cybersecurity issues.
But, being statistical, they can be applied "black box", in particular by deployers.
Rely on defintions of accuracy.

SC, Standardisation, and the EU AI Act: IS 24029–2 [ISO23]

AI — Assessment of the robustness of neural networks — Part 2: Methodology for the use of formal methods

This standard really is about "neural networks". Can look at Stability, Sensitivity, Relevance and Reachability.
Basic methodologies are MILP (Mixed Integer–Linear Programming), SAT and SMT, e.g. [KBD+17].
MILP (and some SMT) assume $\mathbf{R}$. Traditionally computer programs have been $\mathbf{R}_{IEEE-754}$, generally in 64-bit. Most people ignore this difference (at least most of the time).
But much "AI" runs in lower precision, e,g. 754/16 bit (semantics are weak), `bfloat` (no formal semantics), or even 8-bit (semantics coming in IEEE-P3109 [IEE25, but several alternatives]) or less.
Maybe the answer here has to be a variant of bit-blasting. **Very little work here**

SC, ~~Standardisation,~~ and the EU AI Act: "external"

My terminology — basically accept that AI tools may not be robust (artificial bullshit!) and build around them.
Note this is not "guardrails" (attempts to prevent AI from generating certain things): these have well-known weaknesses, and all cybersecurity workers are well aware of weaknesses of black-listing/sanitisation.
Rather, can we build a symbolic system on top of AB.
Look at AlphaProof's success in solving Mathematics Olympiad problems [Cas24].

> *Indeed, but, despite the Alpha... name, this is NOT just a trained version of their LLM technology. As I understand it, they have a special Math Engine (essentially computer algebra) to check the solutions. Which is why it can do the algebra problems, but stalls on combinatorics. [JHD]*

SC, Standardisation, and the EU AI Act: Conclusions

Note again that these are personal observations.
There is a great deal to be done in making AI/AB robust, and this will not come from improving the current LLM-style technology.
We need two approaches.

Internal Better checks on the AI systems, such as looking for large values [SZS+13, etc.].

More Research needed, e.g. [KBD+17] is limited to ReLU networks, etc.

External Building true symbolic models to produce "neurosymbolic reasoning" — a phrase I see more of, but have no idea how/if it can be made standard.

What would a good combination look like?


Alcuin of York.
Propositiones ad acuendos iuuenes.
Manuscript (approximate date), 796.
URL: https://www.math.muni.cz/~sisma/alcuin/alcuin_latinsky.pdf.


R.J. Anderson and R.M. Needham.
Programming Satan's Computer.
In *Computer Science Today*, volume 1000 of *Springer Lecture Notes in Computer Science*, pages 426–440, 1995.
URL: https://link.springer.com/chapter/10.1007/BFb0015258.


D. Castelvecchi.
DeepMind hits milestone in solving maths problems — AI's next grand challenge.
https://www.nature.com/articles/d41586-024-02441-2, 2024.


H.G. Frankfurt.

On Bullshit.
Princeton University Press, 2005.


W.T. Gowers.
It is well known that ChatGPT is bad at problems to do with crossing a river with animals.
https://twitter.com/wtgowers/status/1804565549789135256, 2024.


IEEE.
IEEE Working Group P3109 Interim Report on Binary Floating-point Formats for Machine Learning (Version 3.0.3).
https://github.com/P3109/Public/blob/main/IEEE%20WG%20P3109%20Interim%20Report%20v3.pdf, 2025.


ISO/IEC SC42.
ISO/IEC TR 24029-1:2021 Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview.

ISO/IEC, 2021.


ISO/IEC SC42.
ISO/IEC TR 24029-2:2023 Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 2: Methodology for the use of formal methods.
ISO/IEC, 2023.


ISO/IEC SC42.
ISO/IEC TR 24029-3:2025 Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 3: Methodology for the use of statistical methods.
ISO/IEC Committee Draft, 2025.


G. Katz, C.W. Barrett, D.L. Dill, K. Julian, and M.J. Kochenderfer.
Reluplex: An Efficient SMT Solver for Verifying Deep Neural Networks.
In *Proceedings International Conference on Computer Aided Verification*, pages 97–117, 2017.


C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus.
Intriguing properties of neural networks.
https://arxiv.org/abs/1312.6199, 2013.