ZSY1 Definice AI

středa 14. února 2024 12:54

Definice AI

Definice umělé inteligence

Umělá inteligence - Artificial intelligece (AI) - je schopnost strojů napodobovat lidské schopnosti, jako je uvažování, učení se, plánování nebo kreativita.

Umělá inteligence umožňuje technickým systémům reagovat na vněmy z jejich prostředí, řešit problémy a dosahovat určitých cílů. Zabudovaný počítač přijímá data - která byla již připravena, nebo jsou sbírána pomocí vlastních sensorů a kamer - ty následně vyhodnotí a reaguje na ně.

Systémy umělé inteligence jsou schopné pracovat samostatně a také měnit a přizpůsobovat své jednání na základě vyhodnocení efektů předchozích akcí.

S ohledem na blížící se volby do Evropského parlamentu je tak možné očekávat, že by schválený návrh mohl vstoupit v platnost (a tedy být zveřejněn v Úředním věstníku EU) ještě před létem 2024, nicméně se v tuto chvíli stále jedná pouze o náš hrubý odhad.

Co se změnilo?:

Ustanovení GPAI:

▪ Definice: „univerzálním modelem umělé inteligence" se rozumí model umělé inteligence, včetně toho, když je trénován s velkým množstvím dat pomocí mechanismu učení bez učitele, který vykazuje významnou obecnost a je schopen kompetentně provádět širokou škálu různých úkolů bez ohledu na způsob, jakým je model uveden na trh a který lze integrovat do různých navazujících systémů nebo aplikací.

o Bod odůvodnění objasňující, že Komise může aktualizovat technické prvky definice modelu GPAI.

▪ Systém GPAI je definován jako systém umělé inteligence, který je založen na modelu umělé inteligence pro obecné účely, který má schopnost sloužit různým účelům, a to jak pro přímé použití, tak pro integraci do jiných systémů umělé inteligence.

▪ „následným poskytovatelem systému umělé inteligence" se rozumí jakýkoli poskytovatel systému umělé inteligence, včetně systému GPAI, který integruje model umělé inteligence pro obecné účely.

▪ Systémové riziko je rozvedeno v bodech odůvodnění jako negativní dopady na závažné havárie, narušení v kritických sektorech, ohrožení veřejného zdraví a bezpečnosti, jakož i dopady na demokratické procesy, veřejnou a ekonomickou bezpečnost. že tato rizika se mohou zvyšovat se schopnostmi a dosahem modelů umělé inteligence a mohou se projevovat v průběhu životního cyklu modelu. V bodech odůvodnění jsou uvedeny různé faktory ovlivňující tato rizika, jako je zneužití,

spolehlivost modelu, spravedlnost, bezpečnost, autonomie, přístup k nástrojům, distribuční strategie a potenciální odstranění záruk.

Pravidla klasifikace:

Klasifikace modelů GPAI jako modelů GPAI se systémovým rizikem:

• Model GPAI je klasifikován jako model GPAI se systémovým rizikem, pokud splňuje některé z kritérií:

o Schopnosti s vysokým dopadem hodnocené na základě vhodných nástrojů a metodik

o Rozhodnutí úřadu AI nebo kvalifikované upozornění vědeckého panelu, že model GPAI má schopnosti dopadu.

• Předpokládá se, že model GPAI má velký dopad, pokud je jeho kumulativní množství výpočtů použitých při tréninku větší než 10^25 FLOP

• Komise může změnit prahové hodnoty a kritéria prostřednictvím aktů v přenesené pravomoci.

Postup:

• Pokud model GPAI splňuje požadavky na klasifikaci jako systémové riziko, musí to poskytovatel oznámit Komisi bez prodlení/maximálně dvou týdnů spolu s nezbytnými informacemi. Pokud se to Komise dozví, aniž by o tom byla informována, může přistoupit k označení modelu za systémové riziko.

• Poskytovatelé modelu systémového rizika GPAI mohou v procesu oznamování Komisi předložit argumenty vyvracející označení.

• Komise může označit model GPAI za systémové riziko, a to z moci úřední nebo na základě kvalifikovaných výstrah vědeckého panelu na základě kritérií stanovených v nové příloze. Komise je zmocněna upřesňovat a aktualizovat kritéria této přílohy prostřednictvím aktů v přenesené pravomoci.

• Komise bude udržovat a zveřejňovat seznam modelů GPAI se systémovým rizikem při respektování práv duševního vlastnictví, obchodních tajemství a důvěrných obchodních informací.

Povinnosti poskytovatelů modelu GPAI:

• Vypracovat technickou dokumentaci modelu včetně školení a procesu testování minimálně obsahující informace v nové příloze o technické dokumentaci modelů GPAI - připravená na vyžádání Úřadu AI a NCA.

• Vypracovávat a udržovat aktuální informace a poskytovat dokumentaci poskytovatelům systémů umělé inteligence, kteří mají v úmyslu integrovat modely GPAI do svých systémů. Dokumentace musí:

o Umožněte poskytovatelům systémů umělé inteligence dobře porozumět možnostem a omezením modelu GPAI a splnit své závazky AIA.

o obsahovat minimálně podrobnosti z přílohy XY

o Zavést politiku respektující autorské právo Unie, zejména výhradu práv vyjádřenou v čl. 4 odst. 3 směrnice o autorském právu o dolování textů a dat.

o Vypracovat a zveřejnit dostatečně podrobné shrnutí obsahu školení podle šablony poskytnuté kanceláří AI.

• Poskytovatelé modelů GPAI se mohou při plnění výše uvedených povinností spoléhat na kodexy správné praxe, dokud nebude vydána harmonizovaná norma. Poskytovatelé, kteří nedodržují schválený kodex, musí prokázat alternativní způsoby shody.

• Povinnosti se nevztahují na poskytovatele předem vyškolených modelů zpřístupněných na základě licence, která umožňuje veřejný přístup, použití, úpravy a distribuci a zahrnuje váhy a informace o architektuře modelu a použití modelu.

Autorská práva:

• Body odůvodnění uznávají, že techniky dolování textu a dat používané v kontextu modelu GPAI mohou zahrnovat chráněný obsah, a body odůvodnění vysvětlují podmínky, za kterých je takové dolování povoleno podle směrnice (EU) 2019/790. Dále zdůrazňuje, že je nutné, aby poskytovatelé modelů umělé inteligence pro všeobecné účely dodržovali příslušné zákony o autorských právech a získali oprávnění od držitelů práv pro činnosti dolování textů a dat. Držitelé práv mohou také použít metodu „opt-out". Poskytovatelé jsou povinni mít zásady respektující právo Unie o autorském právu a souvisejících právech, konkrétně identifikující a respektující práva vyhrazená držiteli práv. Tato povinnost platí bez ohledu na to, kde probíhají akty související s autorským právem, na nichž je školení těchto modelů založeno.

• Poskytovatelé budou také muset vypracovat a zveřejnit dostatečně podrobné shrnutí obsahu školicích údajů, které by mělo být komplexní, aniž by bylo technicky podrobné, aby umožňovalo určitou volnost, pokud jde o obchodní tajemství a ochranu důvěrných informací. Shrnutí by však mělo stranám s oprávněným zájmem stále umožňovat vymáhání jejich nároků na autorská práva – například uvedením hlavních souborů sběru dat, jako jsou veřejné databáze nebo archivy.

Povinnosti pro poskytovatele GPAI se systémovým rizikem:

• Kromě základních povinností pro modely GPAI jsou další požadavky na systémové riziko:

o a) provést vyhodnocení modelu pomocí standardizovaných protokolů a nástrojů odrážejících současný stav techniky;

o b) posuzovat a zmírňovat možná systémová rizika na úrovni Unie, včetně jejich zdrojů, která mohou pocházet z vývoje, uvádění na trh, uvádění do provozu nebo používání obecných modelů umělé inteligence se systémovým rizikem;

o c) sledovat, dokumentovat a bez zbytečného prodlení podávat Komisi a případně příslušným vnitrostátním orgánům příslušné informace o závažných incidentech a možných nápravných opatřeních k jejich řešení;

o d) Provádět a dokumentovat nepříznivé testování modelu vedle nebo jako součást plnění povinnosti podle čl. C odst. 1 písm. a) s cílem identifikovat a zmírnit systémová rizika;

o e) zajistit přiměřenou úroveň ochrany kybernetické bezpečnosti pro obecný model umělé inteligence se systémovým rizikem a fyzickou infrastrukturu modelu;

o f) sledovat dokumenty a podávat zprávy o známé nebo odhadované spotřebě energie modelu; v případě, že to není známo nebo pokud není k dispozici žádný standard, může to být založeno na informacích o použitých výpočetních zdrojích.

• 2. Poskytovatelé obecných modelů umělé inteligence se systémovým rizikem se mohou spoléhat na kodexy správné praxe ve smyslu článku (Kodex správné praxe) k prokázání souladu s povinnostmi v odstavci 1, dokud nebude zveřejněna harmonizovaná norma. Shoda s evropskou harmonizovanou normou poskytuje poskytovatelům předpoklad shody. Poskytovatelé univerzálních modelů umělé inteligence se systémovými riziky, kteří nedodržují schválený kodex správné praxe, prokáží ke schválení Komisi alternativní přiměřené způsoby shody.

• 3. Je-li to nezbytně nutné pro účely splnění povinností stanovených v tomto článku, obchodní tajemství bude zachováno v souladu se směrnicí (EU) 2016/943 a bude zveřejněno pouze za předpokladu, že budou předem přijata všechna konkrétní nezbytná opatření. zachovat jejich důvěrnost, zejména pokud jde o třetí osoby.

• 4. Výjimka stanovená v čl. C odst. 4 se nevztahuje na modely umělé inteligence pro obecné účely se systémovým rizikem.

Článek o zásadách správné praxe

• Ai Office bude podporovat vypracování kodexů s ohledem na mezinárodní přístupy.

• Kodexy by měly pokrývat povinnosti stanovené pro poskytovatele v článcích.

• Kanceláří AI může vyzvat zúčastněné strany k účasti na vypracování kodexů správné praxe.

• Kanceláří a rada AI se zaměří na zajištění jasných cílů a KPI, které jsou načrtnuty a nezbytné.

• Komise zajistí, aby se účastníci pravidelně přihlašovali, zatímco úřad a správní rada budou monitorovat a hodnotit. Komise může použít prováděcí akt ke schválení kodexů a získání obecné platnosti v EU.

• Kodexy správné praxe musí být vypracovány nejpozději devět měsíců před vstupem v platnost. Pokud tak není učiněno včas, může Komise zasáhnout prováděcím aktem.

Budou také vytvořeny tři nové přílohy:

• technická dokumentace pro poskytovatele modelů GPAI,

• transparentní informace týkající se technické dokumentace potřebné pro následné poskytovatele,

• stanovení schopností modelu GPAI.

Harmonizované standardy a standardizace

• Nový dodatek k článku 40, který upřesňuje, že požadavky na standardizaci mohou rovněž vyžadovat výstupy týkající se procesu podávání zpráv a dokumentace pro výkon zdrojů systémů umělé inteligence, jako je snížení spotřeby energie a vývoj modelů GPAI v oblasti energetické účinnosti.

• Dva roky po podání žádosti a každé čtyři roky poté Komise předloží zprávu o přezkumu pokroku v oblasti normalizačních výstupů v oblasti energeticky účinného vývoje modelů GPAI a posoudí potřebu dalších opatření.

GPAI Provisions:

- Definition: 'general-purpose AI model' means an AI model, including when trained with a large amount of data using self-supervision at scale, that

displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is released on the market and that can be integrated into a variety of downstream systems or application.
    - Recital clarifying that the Commission can update the technical elements of the GPAI model definition.

- GPAI System is defined as an AI system that is based on a general-purpose AI model that has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems.

- 'downstream provider of an AI system' means any provider of an AI system, including a GPAI system, who integrates a general-purpose AI model.

- Systemic Risk is elaborated in the recitals as negative effects on major accidents, disruptions in critical sectors, threats to public health and safety, as well as impacts on democratic processes, public and economic security. that these risks may increase with the capabilities and reach of the AI models, and they can manifest throughout the model's lifecycle. The recitals list out various factors influencing these risks, such as misuse, model reliability, fairness, security, autonomy, access to tools, distribution strategies, and potential removal of safeguards.

Classification Rules:

Classification of GPAI Models as GPAI Models with Systemic Risk:

- GPAI model is classified as GPAI model with systemic risk if it meets any of the criteria:
    - High impact capabilities evaluated based on appropriate tools and methodologies
    - Decision of the AI Office or a qualified alert of the scientific panel that a GPAI model has capabilities of impact.
- GPAI Model is presumed to be high impact if its cumulative amount of compute used in training is greater than 10^25 FLOP
- The Commission may amend the thresholds and criteria via delegated acts.

Procedure:

- When the GPAI model meets the requirements to be classified as a systemic risk the provider must notify the Commission without delay/max two weeks, along with the information necessary. If the Commission becomes aware without having been notified it may proceed with designating the model a systemic risk.
- Providers of the GPAI model of systemic risk may, in the process of notifying the Commission, present arguments rebutting the designation.

- The Commission may designate the GPAI model as systemic risk, ex officio, or following qualified alerts of the scientific panel based on criteria established in a new Annex. Which the Commission is empowered to specify and update the criteria of this annex via delegated acts.
- The Commission shall maintain and publish a list of GPAI models with systemic risk while respecting IP rights, trade secrets, and confidential business information.

Obligations for Providers of the GPAI Model:

- Draw up technical documentation of the model including training and testing process at a minimum containing the info in a new Annex on technical documentation of GPAI models- ready for the request of the AI Office and NCAs.
- Draw up and keep up-to-date information and provide documentation to providers of AI systems who intend to integrate GPAI models into their systems. The documentation shall:
    - Enable AI system providers to have a good understanding of the capabilities and limitations of the GPAI model and to comply with their AIA obligations.
    - Contain at minimum details from Annex XY
    - Establish a policy to respect Union copyright law in particular the reservation of rights expressed under Art 4.3 of the Copyright Directive on text and data mining.
    - Draw up and make public a sufficiently detailed summary about the content used for training according to the template provided by the AI office.
- Providers of GPAI models may rely on codes of practice to comply with the above obligations until a harmonised standard is released. Providers not adhering to an approved code must demonstrate alternative means of compliance.
- Obligations shall not apply to providers of pre-trained models made accessible under a licence that allows for public access, usage, modification and distribution and includes the weights and information on model architecture and model usage.

Copyright:

- The recitals acknowledge that the text and data mining techniques used in a GPAI model context may involve protected content, and the recitals explain the conditions under which such mining is allowed under Directive (EU) 2019/790. Furthermore, it stresses the need for providers of general-purpose AI models to comply with relevant copyright laws and obtain authorization from rights holders for text and data mining activities. Rightsholders may use the 'opt-out' method as well. Providers are required to have policies respecting Union law on copyright and related rights, specifically identifying and

respecting rights reserved by rightsholders. This obligation applies regardless of where the copyright-relevant acts underpinning the training of these models take place.

- Providers will also need to draw up and publish a sufficiently detailed summary of the content of the training data, which should be comprehensive in scope without being technically detailed to allow for some leeway in respect for trade secrets and confidential information protection. But the summary should still allow for parties with legitimate interest to enforce their copyright claims- so for example listing the main data collection sets such as public databases or archives.

Obligations for providers of GPAI with systemic risk:

- In addition to the base obligations for GPAI models, additional requirements for systemic risk are:
  - a) perform the model evaluation by standardized protocols and tools reflecting the state of the art;
  - b) assess and mitigate possible systemic risks at the Union level, including their sources, that may stem from the development, placing on the market, putting into service or use of general-purpose AI models with systemic risk;
  - c) keep track of, document and report without undue delay to the Commission and, as appropriate, to national competent authorities, relevant information about serious incidents and possible corrective measures to address them;
  - d) Conduct and document adversarial testing of the model in addition to or as part of fulfilling the obligation in Article C (1)(a), with a view to identify and mitigate systemic risks;
  - e) ensure an adequate level of cybersecurity protection for the general purpose AI model with systemic risk and the physical infrastructure of the model;
  - f) keep track of documents and report about known or estimated energy consumption of the model; in case not known or as long as no standard is available, this could be based on information about computational resources used.

- 2. Providers of general purpose AI models with systemic risk may rely on codes of practice within the meaning of Article (Code of Practice) to demonstrate compliance with the obligations in paragraph 1, until a harmonized standard is published. Compliance with a European harmonized standard grants providers the presumption of conformity. Providers of general-

purpose AI models with systemic risks who do not adhere to an approved code of practice shall demonstrate alternative adequate means of compliance for approval of the Commission.

- 3. Where strictly necessary for the purposes of complying with the obligations laid down in this article, trade secrets shall be preserved in accordance with Directive (EU) 2016/943 and shall only be disclosed provided that all specific necessary measures are taken in advance to preserve their confidentiality, in particular concerning third parties.

- 4. The exemption provided for in Article C(4) shall not apply to general-purpose AI models with systemic risk

Code of Practice Article

- The Ai Office will encourage codes to be drawn up, considering international approaches.
- The codes should cover the obligations put forward for providers in the articles.
- AI Office may invite stakeholders to participate in the drawing up of codes of practice.
- The AI Office and Board shall aim to ensure clear objectives and KPIs are outlined and necessary.
- The Commission will ensure participants regularly check-in, while the Office and Board will monitor and evaluate. The Commission may use an implementing act to approve the codes and give general validity in the EU.
- Codes of practice shall be drawn up at the latest nine months before entry into application. If this is not done in time the Commission may intervene with an implementing act.

Three new annexes will also be created:

- technical documentation for providers of GPAI models,
- transparency information relating to the technical documentation needed for downstream providers,
- determining the capabilities of a GPAI Model.

Harmonised Standards and Standardisation Deliverable-

- New addition to Article 40 specifying that standardisation requests may also ask for deliverables on reporting and documentation process for AI systems

resource performance such as for reduction of energy and on energy efficiency development of GPAI models..

- Two years after application and every four after the Commission shall submit a report on the review of the progress on standardisation deliverables on energy efficient development of GPAI models and assess need for further action