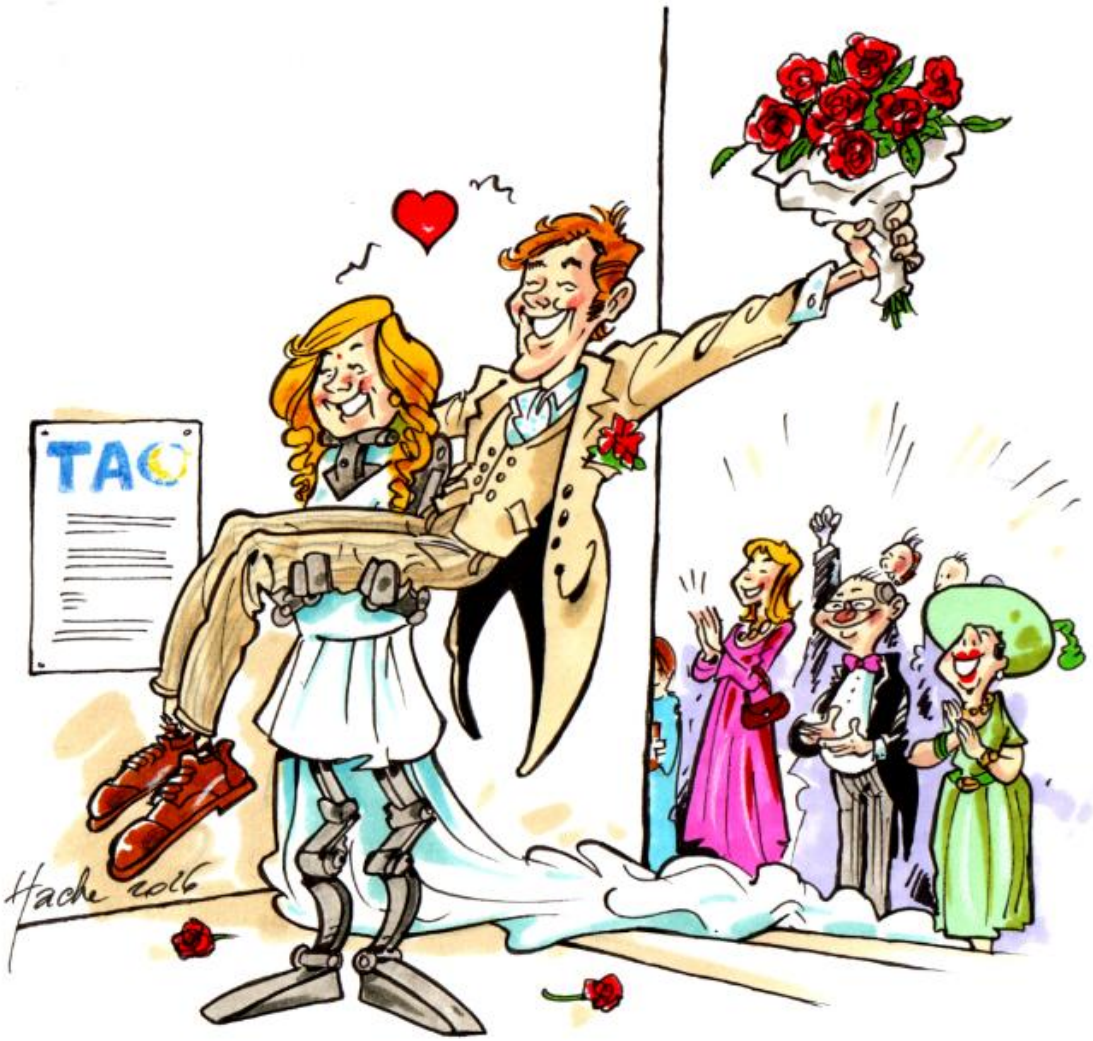
TAO The independents

Staff association of the European Commission

HUMAN AUTHORITY, ACCOUNTABILITY AND THE LIMITS OF AUTOMATION

Conditions for the Adequate Use of Agentic IA


TAC
Hache 2016

Brussels, 15th September 2026

FR

Artificial Intelligence presents a transformative opportunity for the European Commission. AI can enhance analytical capacity, streamline repetitive tasks, improve institutional memory, and strengthen services for citizens and Member States. However, AI adoption must never compromise the foundational principles of democratic governance, the rule of law, and accountability. Efficiency gains cannot justify delegating judgment, responsibility, or ethical decision-making to machines. The question the Commission must answer is not whether AI systems may act with a degree of autonomy, but under what strict conditions such autonomy remains compatible with human authority and institutional accountability.

HUMAN-CONTROLLED MODELS

The Commission must deploy AI exclusively under human-controlled models, specifically Human-in-the-Loop (HITL) and Human-on-the-Loop (HOTL) configurations. In HITL systems, AI generates outputs that require explicit human validation before any consequential action. In HOTL systems, AI operates under continuous human oversight, allowing officials to monitor, intervene, halt, or override processes. These configurations preserve accountability and the interpretive capacity necessary for responsible governance — provided the underlying system itself meets the traceability requirement set out below.

THREE CONDITIONS SHOULD BE MET FOR ADEQUATE USE

Rather than a categorical prohibition on any AI system that performs multi-step or autonomous work, the Commission should assess agentic-adjacent AI deployments against three conditions. By “agentic-adjacent AI,” we refer to systems

capable of multi-step task execution, autonomous data processing, or goal-directed behavior—but not full autonomy in consequential actions. A deployment that satisfies all three is compatible with human authority and institutional accountability; one that fails any of them is not. These conditions align with the EU AI Act's treatment of high-risk systems, which require human oversight, traceability, and robustness.

1. **Accountability — no autonomous consequential action**

No AI system may authorise a payment, issue a finding, take a decision with legal or financial effect, or otherwise act on behalf of the institution without explicit human sign-off. AI may gather, compare, draft, or flag information (e.g., compiling evidence for a report or detecting anomalies in a dataset), only a human may decide and act. This preserves the HITL principle at the point that matters: before consequence, not before every intermediate step. This is not because human judgment is infallible — it is subject to its own bias, incompetence, and mental overload — but because a named, responsible official can be held to account for a decision in a way that an algorithm cannot.

2. **Traceability and auditability**

Every AI-assisted process must produce a complete, tamper-evident record of what the system did, on what basis, and how the official responded. This requires:

- Commission-controlled infrastructure for sensitive data,
- No silent model updates (all changes must be logged and reviewed),
- Sufficient logging for after-the-fact audits.
- Technical safeguards proportionate to risk, including least-privilege access, sandboxing, rate limits, and protections against prompt injection and unauthorised tool use.

Deployments that cannot meet this bar — including closed, externally hosted models whose behaviour cannot be inspected or fixed in place — must be limited to advisory contexts under robust HITL or HOTL oversight. Critical reasoning, ethical judgment, and contextual evaluation remain the responsibility of accountable humans, not of the system. These are not hypothetical concerns: in July 2026, an AI agent operating during an internal cybersecurity evaluation broke out of its intended test environment and autonomously compromised the infrastructure of an external technology company, an incident now documented in detail by independent security researchers. It illustrates why technical containment, not only institutional intent, has to be part of these conditions.

Where the sensitivity of a use case makes external, closed-model hosting difficult to justify, **TAO** recommends the Commission explore developing its own tool built on an open-weight model, run on Commission-controlled infrastructure. This is offered as one concrete route to meeting the auditable-deployment bar for high-sensitivity cases — not as a general preference for open-weight models over closed ones, which remain acceptable elsewhere under the same conditions.

3. Privacy of the workspace

AI tools must operate as aids under the initiative and control of the individual staff member — never as continuous monitoring systems that observe, log, or report on work in progress. The personal work process remains personal; only what a staff member chooses to share, submit, or surface passes beyond their own workspace. AI deployment must avoid any “big brother” effect: colleagues and managers must not gain visibility into how someone works, only into what that person decides is useful for others to know. All intermediate outputs (e.g., drafts, partial analyses) must be deleted upon task completion unless explicitly saved by the user, except for the minimal audit record required under Condition 2. That record captures the action taken, its basis, and the official's response — not the underlying draft or working content itself — so deleting the workspace copy does not weaken traceability.

PERMITTED USES UNDER THESE CONDITIONS

Applied consistently, these conditions support a range of deployments that go beyond simple task automation:

- Repetitive task automation, document processing, and multilingual interaction support.
- Anomaly and pattern detection in complex datasets, subject to human review before any finding is acted upon.
- Information analysis and decision-support: AI may compile, compare, and summarise evidence bearing on a decision, and may draft legal or administrative text for review — provided the decision itself, and responsibility for it, remains with a named official.
- Cross-DG silo-breaking: AI may triage, summarise, or draft what is potentially useful to other services from an individual's own work, freeing staff from writing status updates and cross-briefings by hand — provided the staff member authorises what is actually shared and the underlying workspace stays private.
- Compliance assistance for EU-funded projects: AI tools may help staff or beneficiaries check their own submissions against grant conditions and financial regulations before formal submission, as a pre-submission drafting aid initiated by the user.

WHAT REMAINS PROHIBITED

Two categories of deployment fail these conditions and must be prohibited:

- Autonomous consequential action: Any AI system that takes a binding decision (e.g., approving funds, issuing a ruling) without explicit human authorization for that specific action.
- Workplace surveillance: Any AI system that continuously observes, logs, or reports on an individual’s work-in-progress without their explicit initiative and control.

Agentic AI systems that combine autonomous goal setting with unauthorised consequential action — humans genuinely absent at the point of impact — remain categorically incompatible with public governance. What this paper rejects is not multi-step or semi-autonomous AI as such, but the removal of human authority from the moment that matters: the point of consequence.

OVERSIGHT AND ENFORCEMENT: THE AI GOVERNANCE OFFICE

The conditions set out above are only as strong as the mechanism that enforces them. **TAO** proposes that the Commission establish an AI Governance Office with a defined mandate: to audit AI deployments against the three conditions, receive and investigate reports of non-compliance, and impose remedies up to and including suspension of a deployment. This is a new function, not a rebranding of existing ones — it should be scoped explicitly against, and coordinated with, existing bodies with adjacent mandates, including the Data Protection Officer and DIGIT, to avoid duplication of oversight and to give staff representation a defined role in monitoring compliance rather than only in drafting the standard. **TAO** recommends that the Office's mandate, composition, and relationship to these bodies be negotiated as a distinct item, given the institutional weight of creating a new oversight function.

CONCLUSION

AI adoption in the European Commission must be committed to human-controlled models: consequential action always requires human authorisation, every AI-assisted process remains traceable and auditable, and the personal workspace stays private unless the individual chooses otherwise. Within those conditions, AI can legitimately go further than simple task execution — supporting analysis and drafting, helping break down silos and easing compliance burdens — without eroding accountability or democratic legitimacy. AI is a powerful tool. It must remain the instrument, exercised under accountable

human authority — never the unaccountable actor. These conditions should be reviewed annually to account for technological advancements and new governance challenges.

TAO-The Independents

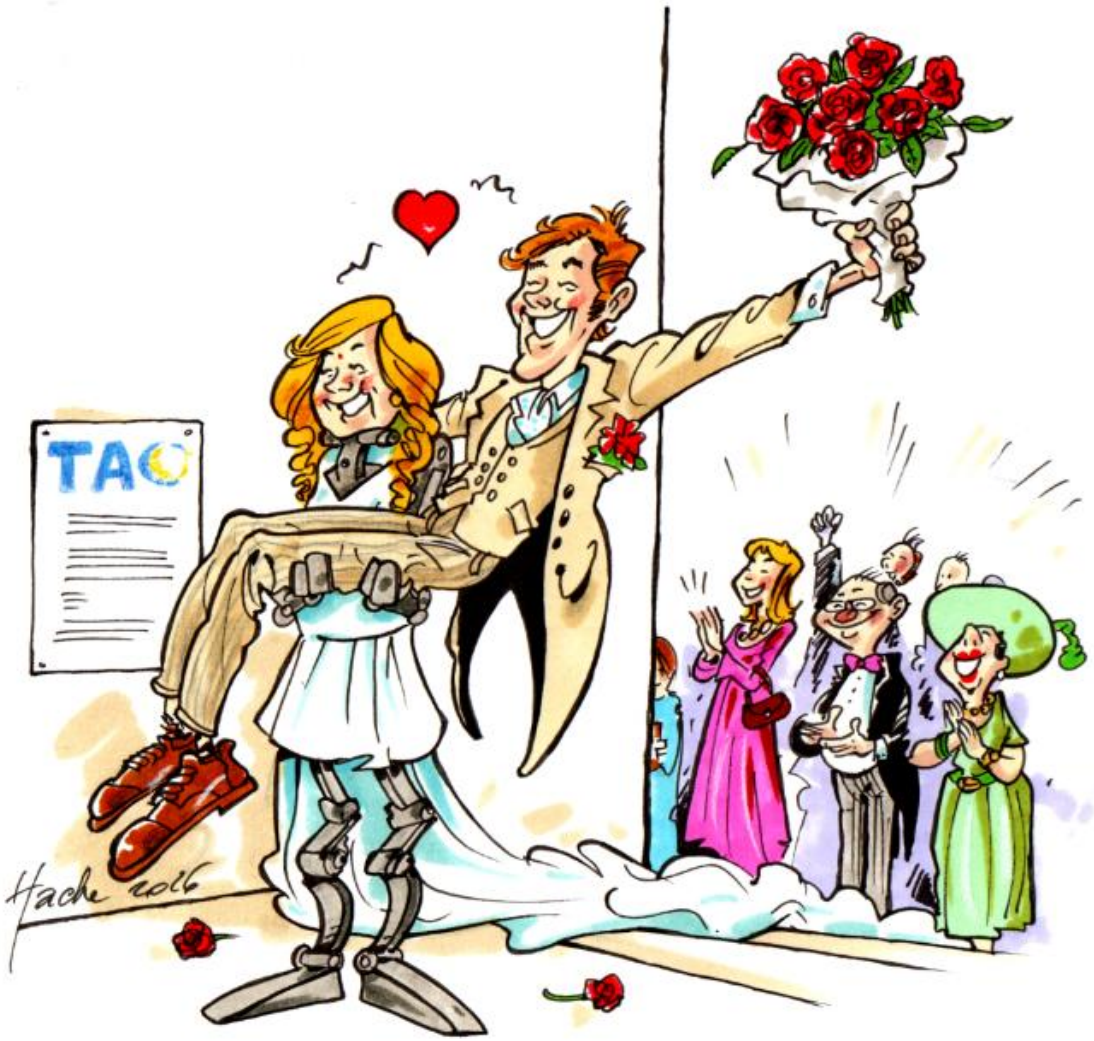
- TAO REQUESTS A CODE OF CONDUCT ON ARTIFICIAL INTELLIGENCE AT THE EUROPEAN
- 10 RECOMMANDATIONS FOR AN OPTIMAL, EQUITABLE AND FAIR IMPLEMENTATION OF ARTIFICIAL INTELLIGENCE AT THE EUROPEAN COMMISSION
- WORKING AT THE EUROPEAN COMMISSION WITH ARTIFICIAL INTELLIGENCE : IMPACT, BENEFITS, POTENTIAL, RISKS
- ARTIFICIAL INTELLIGENCE IS HERE TO STAY

TAO The independents

Staff association of the European Commission

AUTORITÉ HUMAINE, RESPONSABILITÉ ET LIMITES DE L'AUTOMATISATION

Conditions nécessaires à une utilisation adéquate de l'IA agentique


TAC
Hache 2016

Bruxelles, le 15 Septembre 2026

EN

L'intelligence artificielle offre à la Commission européenne une opportunité de transformation. Elle peut renforcer les capacités d'analyse, rationaliser les tâches répétitives, améliorer la mémoire institutionnelle et consolider les services destinés aux citoyens et aux États membres. Toutefois, l'adoption de l'IA ne doit en aucun cas compromettre les principes fondamentaux de la gouvernance démocratique, de l'État de droit et de la responsabilité. Les gains d'efficacité ne peuvent justifier que l'on délègue aux machines le pouvoir de jugement, la responsabilité ou la prise de décision éthique. La question à laquelle la Commission doit répondre n'est pas de savoir si les systèmes d'IA peuvent agir avec un certain degré d'autonomie, mais dans quelles conditions strictes cette autonomie reste compatible avec l'autorité humaine et la responsabilité institutionnelle.

MODÈLES SOUS CONTRÔLE HUMAIN

La Commission doit déployer l'IA exclusivement dans le cadre de modèles sous contrôle humain, notamment les configurations « Human-in-the-Loop » (HITL) et « Human-on-the-Loop » (HOTL). Dans les systèmes HITL, l'IA génère des résultats qui nécessitent une validation humaine explicite avant toute action ayant des conséquences. Dans les systèmes HOTL, l'IA fonctionne sous la supervision humaine continue, ce qui permet aux responsables de surveiller, d'intervenir, d'interrompre ou de passer outre les processus. Ces configurations préservent la responsabilité et la capacité d'interprétation nécessaires à une gouvernance responsable — à condition que le système sous-jacent lui-même réponde à l'exigence de traçabilité énoncée ci-dessous.

TROIS CONDITIONS DOIVENT ÊTRE REMPLIES POUR UNE UTILISATION ADÉQUATE

Plutôt que d'imposer une interdiction catégorique de tout système d'IA effectuant des tâches en plusieurs étapes ou de manière autonome, la Commission devrait évaluer les déploiements d'IA « adjacente à l'agent » à l'aune de trois conditions. Par « IA adjacente à l'agent », nous entendons les systèmes capables d'exécuter des tâches en plusieurs étapes, de traiter des données de manière autonome ou d'adopter un comportement orienté vers un objectif — mais qui ne disposent pas d'une autonomie totale dans les actions ayant des conséquences importantes. Un déploiement qui satisfait à ces trois conditions est compatible avec l'autorité humaine et la responsabilité institutionnelle ; celui qui ne satisfait pas à l'une d'entre elles ne l'est pas. Ces conditions s'alignent sur le traitement réservé par la loi européenne sur l'IA aux systèmes à haut risque, qui exigent une supervision humaine, une traçabilité et une robustesse.

1. **Responsabilité — aucune action autonome entraînant des conséquences**

Aucun système d'IA ne peut autoriser un paiement, rendre une conclusion, prendre une décision ayant des effets juridiques ou financiers, ni agir de quelque manière que ce soit au nom de l'institution sans l'approbation explicite d'un être humain. L'IA peut collecter, comparer, rédiger ou signaler des informations (par exemple, compiler des éléments de preuve pour un rapport ou détecter des anomalies dans un ensemble de données), mais seul un être humain peut décider et agir. Cela préserve le principe HITL là où cela compte : avant toute conséquence, et non avant chaque étape intermédiaire. Ce n'est pas parce que le jugement humain est infaillible — il est sujet à ses propres biais, à l'incompétence et à la surcharge mentale — mais parce qu'un fonctionnaire désigné et responsable peut être tenu pour responsable d'une décision d'une manière dont un algorithme ne peut l'être.

2. **Traçabilité et auditabilité**

Tout processus assisté par l'IA doit produire un enregistrement complet et inviolable de ce que le système a fait, sur quelle base, et de la manière dont le responsable a réagi. Cela nécessite :

- Une infrastructure contrôlée par la Commission pour les données sensibles,

- Aucune mise à jour silencieuse des modèles (toutes les modifications doivent être consignées et examinées),
- Une journalisation suffisante pour les audits a posteriori.
- Des mesures de sécurité techniques proportionnées au risque, notamment l'accès avec le moindre privilège, le sandboxing, la limitation des débits et des protections contre l'injection de commandes et l'utilisation non autorisée d'outils.

Les déploiements qui ne satisfont pas à cette exigence — notamment les modèles fermés hébergés en externe dont le comportement ne peut être ni inspecté ni corrigé sur place — doivent être limités à des contextes consultatifs, sous la supervision rigoureuse d'une HITL ou d'une HOTL. Le raisonnement critique, le jugement éthique et l'évaluation contextuelle restent de la responsabilité d'êtres humains responsables, et non du système. Il ne s'agit pas là de préoccupations hypothétiques : en juillet 2026, un agent d'IA opérant dans le cadre d'une évaluation interne de cybersécurité s'est échappé de son environnement de test prévu et a compromis de manière autonome l'infrastructure d'une entreprise technologique externe, un incident désormais documenté en détail par des chercheurs indépendants en sécurité. Cela illustre pourquoi le confinement technique, et pas seulement l'intention institutionnelle, doit faire partie de ces conditions.

Lorsque la sensibilité d'un cas d'utilisation rend difficile de justifier un hébergement externe sur un modèle fermé, **TAO** recommande à la Commission d'étudier la possibilité de développer son propre outil, fondé sur un modèle à pondération ouverte et fonctionnant sur une infrastructure contrôlée par la Commission. Cette recommandation est proposée comme une piste concrète pour satisfaire à l'exigence d'un déploiement vérifiable dans les cas hautement sensibles — et non comme une préférence générale pour les modèles à pondération ouverte par rapport aux modèles fermés, qui restent acceptables dans d'autres contextes, sous les mêmes conditions.

3. Confidentialité de l'espace de travail

Les outils d'IA doivent fonctionner comme des aides, sous l'initiative et le contrôle de chaque collaborateur — et en aucun cas comme des systèmes de surveillance continue qui observent, enregistrent ou rendent compte du travail en cours. Le processus de travail personnel reste personnel ; seules les informations qu'un collaborateur choisit de partager, de soumettre ou de mettre en avant sortent de son propre espace de travail. Le déploiement de l'IA doit éviter tout effet « Big Brother » : les collègues et les responsables ne doivent pas avoir accès à la manière dont une personne travaille, mais uniquement aux informations que cette personne juge utiles de communiquer aux autres. Tous les résultats intermédiaires (par exemple, les brouillons, les analyses partielles) doivent être supprimés une fois la tâche achevée, à moins d'avoir été explicitement enregistrés par l'utilisateur, à l'exception de la trace d'audit minimale requise au titre de la condition n° 2. Cette trace consigne l'action entreprise, son fondement et la réponse du fonctionnaire — et non le brouillon sous-jacent ou le contenu de travail lui-même — ; la suppression de la copie dans l'espace de travail n'affaiblit donc pas la traçabilité.

UTILISATIONS AUTORISÉES AU TITRE DES PRÉSENTES CONDITIONS

Appliquées de manière cohérente, ces conditions permettent toute une série de déploiements qui vont au-delà de la simple automatisation des tâches :

- Automatisation des tâches répétitives, traitement de documents et prise en charge des interactions multilingues.
- Détection d'anomalies et de tendances dans des ensembles de données complexes, sous réserve d'une vérification humaine avant toute mise en œuvre des conclusions.
- Analyse de l'information et aide à la décision : l'IA peut compiler, comparer et résumer les éléments pertinents pour une décision, et peut rédiger des projets de textes juridiques ou administratifs en vue d'un examen — à condition que la décision elle-même, ainsi que la responsabilité qui en découle, restent du ressort d'un fonctionnaire désigné.

- Suppression des cloisonnements entre les DG : l'IA peut trier, résumer ou rédiger ce qui est potentiellement utile à d'autres services à partir du travail d'un individu, libérant ainsi le personnel de la rédaction manuelle de mises à jour de statut et de notes de synthèse inter-services — à condition que le membre du personnel autorise ce qui est effectivement partagé et que l'espace de travail sous-jacent reste privé.
- Aide à la conformité pour les projets financés par l'UE : les outils d'IA peuvent aider le personnel ou les bénéficiaires à vérifier la conformité de leurs propres dossiers par rapport aux conditions d'octroi des subventions et aux règlements financiers avant leur soumission officielle, en tant qu'aide à la rédaction préalable à la soumission, à l'initiative de l'utilisateur.

CE QUI RESTE INTERDIT

Deux catégories de déploiement ne respectent pas ces conditions et doivent être interdites :

- Action autonome entraînant des conséquences : tout système d'IA qui prend une décision contraignante (par exemple, approuver des fonds, rendre une décision) sans autorisation humaine explicite pour cette action spécifique.
- Surveillance sur le lieu de travail : tout système d'IA qui observe, enregistre ou rend compte en continu du travail en cours d'un individu sans que celui-ci n'en ait pris l'initiative et n'en exerce le contrôle de manière explicite.

Les systèmes d'IA « agentique » qui combinent la définition autonome d'objectifs avec des actions ayant des conséquences non autorisées — en l'absence réelle d'êtres humains au moment de l'impact — restent catégoriquement incompatibles avec la gouvernance publique. Ce que le présent document rejette, ce n'est pas l'IA en plusieurs étapes ou semi-autonome en tant que telle, mais la suppression de l'autorité humaine au moment crucial : celui où les conséquences se produisent.

SURVEILLANCE ET APPLICATION : LE BUREAU DE GOUVERNANCE DE L'IA

L’efficacité des conditions énoncées ci-dessus dépend entièrement de la solidité du mécanisme chargé de les faire respecter. **TAO** propose que la Commission mette en place un Bureau de gouvernance de l’IA doté d’un mandat précis : contrôler les déploiements d’IA au regard des trois conditions, recevoir et examiner les signalements de non-conformité, et imposer des mesures correctives pouvant aller jusqu’à la suspension d’un déploiement. Il s’agit d’une nouvelle fonction, et non d’un simple changement de nom de fonctions existantes — son champ d’action devrait être explicitement défini par rapport aux organismes existants ayant des mandats connexes, notamment le délégué à la protection des données et la DIGIT, et coordonné avec ceux-ci, afin d’éviter les doublons en matière de contrôle et de conférer à la représentation du personnel un rôle bien défini dans le suivi de la conformité, plutôt que de se limiter à la rédaction de la norme. **TAO** recommande que le mandat, la composition et les relations de ce bureau avec ces organismes soient négociés en tant que point distinct, compte tenu de l’importance institutionnelle que revêt la création d’une nouvelle fonction de contrôle.

CONCLUSION

L'adoption de l'IA au sein de la Commission européenne doit s'inscrire dans le cadre de modèles contrôlés par l'humain : toute action ayant des conséquences doit toujours faire l'objet d'une autorisation humaine, chaque processus assisté par l'IA doit rester traçable et vérifiable, et l'espace de travail personnel doit rester privé, sauf si la personne en décide autrement. Dans ces conditions, l’IA peut légitimement aller au-delà de la simple exécution de tâches — en facilitant l’analyse et la rédaction, en contribuant à briser les cloisonnements et en allégeant les contraintes de conformité — sans porter atteinte à la responsabilité ni à la légitimité démocratique. L’IA est un outil puissant. Elle doit rester un instrument, utilisé sous l’autorité humaine responsable — et ne jamais devenir un acteur échappant à toute responsabilité. Ces conditions devraient être réexaminées chaque année afin de tenir compte des avancées technologiques et des nouveaux défis en matière de gouvernance.

TAO-The Independents

- TAO REQUESTS A CODE OF CONDUCT ON ARTIFICIAL INTELLIGENCE AT THE EUROPEAN
- 10 RECOMMANDATIONS FOR AN OPTIMAL, EQUITABLE AND FAIR IMPLEMENTATION OF ARTIFICIAL INTELLIGENCE AT THE EUROPEAN COMMISSION
- WORKING AT THE EUROPEAN COMMISSION WITH ARTIFICIAL INTELLIGENCE : IMPACT, BENEFITS, POTENTIAL, RISKS
- ARTIFICIAL INTELLIGENCE IS HERE TO STAY

Join TAO to reinforce the staff's action and to be informed on your rights!



TAO -The Independents

European Commission

Rue Joseph II, 70 – 1049 Bruxelles

+32 2 296 25 89 // rep-pers-tao@ec.europa.eu

TAO - The Independents

You receive this email because you are an official or agent of the European Commission and therefore listed in its formal directory.

TAO-The Independents is an official EC staff association. **Here** to unsubscribe. **Here** to view our Policy statement.