

DE: Richtlinien für den Einsatz von KI bei foraus

foraus erlaubt den Einsatz von KI-Tools in der Content-Produktion (Blogs, Publikationen, etc.), um Effizienz zu steigern, unter strengen Bedingungen zur Wahrung von Glaubwürdigkeit, Transparenz und Unabhängigkeit als Aussenpolitik-Think-Tank. Diese Guidelines orientieren sich an Bundesleitlinien (Transparenz, Nicht-Diskriminierung, Verantwortung), EU-Ethics Guidelines und foraus-eigenen Policy-Papern zu KI-Governance. Sie gelten für alle Mitarbeitenden und Freiwilligen; Verantwortung liegt bei Lead-Autor:in.

Allgemeine Prinzipien

KI-Nutzung ist erlaubt, wenn sie die Qualität steigert und Risiken minimiert. Kernprinzipien:

- Menschliche Kontrolle: KI unterstützt, ersetzt aber keine Expertise; finale Inhalte werden immer von Menschen geprüft und genehmigt.
- Transparenz: Jede KI-Nutzung zur Kreierung von Inhalt wird offengelegt (z.B. Fussnote in Publikationen).
- Genauigkeit und Faktenprüfung: KI-Outputs werden faktengeprüft gegen Primärquellen; keine Verbreitung ungetesteter Infos.
- Ethik und Fairness: Vermeidung von Bias, Diskriminierung oder Fehlinfos; Einhaltung DSGVO (Datenschutz).
- Nachhaltigkeit: Begrenzter Einsatz, um Abhängigkeit zu vermeiden; jährliche Review der Guidelines.
- Eigenverantwortung: Die Autor:innen erklären sich bereit, die Eigenständigkeitserklärungsklausel im Ideendraft/*Note de concept* einzuhalten.

Einsatz in Blogs und Publikationen

KI darf für Textgenerierung, Recherche und Editing genutzt werden.

- Dürfen wir KI nutzen? Ja, z.B. für Entwürfe, Zusammenfassungen oder Ideenfindung (Tools: ChatGPT, Claude, Gemini, Perplexity). Immer: „Dieser Text wurde mit KI-Unterstützung erstellt und von [Name] geprüft.“
- Zukunftsbezug und Vorhersagen: Bei spekulativen Szenarien (z.B. „KI in der Aussenpolitik 2030“), diese klar als Hypothesen kennzeichnen; wenn KI etwas Vorhandenes als „neu“ erfindet, streichen oder korrigieren – Faktenprüfung obligatorisch.
- Beispiel-Fussnote: „Entwurf mit GPT-4o unterstützt; finale Version von [Autor] redigiert und faktengeprüft.“

Aspekt	Erlaubt	Nicht erlaubt	Offenlegungspflicht
Text-Entwurf	Ja		Immer Fussnote
Recherche	Ja, Quellen-Generierung	Als einzige Quelle akzeptieren	„KI-unterstützte Recherche“
Übersetzung	Ja, mit Nachprüfung	Ohne menschliche Korrektur	Optional
Zukunfts-Szenarien	Ja, als Hypothese	Als Fakt darstellen	„Szenario basierend auf...“

KI-generierte Bilder und Visuals

KI-Bilder (z.B. Midjourney, DALL-E) sind erlaubt für illustrative Zwecke, nicht für faktenbasierte Darstellungen.

- Regeln: Nur bei Bedarf (z.B. futuristische Szenarien in Blogs); realistische Abbildungen verboten, um Desinformation zu vermeiden.
- Offenlegung: Immer Watermark oder Caption: „KI-generiertes Bild via [Tool]“. Keine Nutzung in offiziellen Reports.
- Alternativen: Bevorzugt lizenzfreie Stockfotos oder eigene Grafiken; KI nur, wenn unverzichtbar.
- Beispiel: In einem Blog zu „KI in der Diplomatie“: Ein abstraktes Bild von Robotern an Verhandlungstischen, mit Caption „KI-generiertes Bild via DALL-E 3“.

Risikomanagement und Schulung

- Risiken prüfen: Vor Nutzung: Bias-Check (z.B. Prompts diversifizieren), Genauigkeits-Test.
- Neuer Eintrag im How to Guide auf foraus-Mitgliederplattform.
- Schulung: Jährlicher Workshop zu KI-Tools und Ethik (inspiriert von Digital Responsibility CH).
- Sanktionen: Eine nicht deklarierte Verwendung künstlicher Intelligenz kann zu zusätzlichen Überarbeitungen während des Bearbeitungsprozesses oder sogar zur Ablehnung der Veröffentlichung führen.

Implementierung und Update

Diese Guidelines treten sofort in Kraft; verantwortlich für das Informieren von Autor:innen: GS Engagement. Für die Umsetzung tragen die Autor:innen und die Reviewer:innen die Verantwortung. Jährliche Evaluation basierend auf Feedback und neuen Regulierungen (z.B.

EU AI Act, Bundesupdate). Für Anpassungen: engagement@foraus.ch. Jährlich geupdatete KI-Guidelines stärken unsere Rolle als vertrauenswürdiger Think Tank.

Ressourcen für Autor:innen

- [Recommendations on the Use of Generative Artificial Intelligence at UZH | University of Zurich](#)
- [AlgorithmWatch's guidelines to use generative AI responsibly](#)
- [Leitlinien zum Umgang mit Künstlicher Intelligenz in Kommunikation und Marketing an der Universität Bern](#)
- [EU AI Act](#)

FR : Directives relatives à l'utilisation de l'IA au foraus

Le foraus autorise l'utilisation d'outils d'Intelligence Artificielle (IA) dans la production de contenu (blogs, publications, etc.) afin d'améliorer l'efficacité, sous réserve de conditions strictes visant à préserver la crédibilité, la transparence et l'indépendance du foraus en tant que groupe de réflexion sur la politique étrangère. Ces directives s'inspirent des lignes directrices fédérales (transparence, non-discrimination, responsabilité), des lignes directrices éthiques de l'UE et des documents de politique du foraus relatifs à la gouvernance de l'IA. Elles s'appliquent à tous les collaborateur·rice·s et bénévoles ; la responsabilité incombe à auteur·e principal·e.

Principes généraux

L'utilisation de l'IA est autorisée lorsqu'elle améliore la qualité et minimise les risques. Principes fondamentaux :

- Contrôle humain : l'IA apporte un soutien, mais ne remplace pas l'expertise ; les contenus finaux sont toujours vérifiés et approuvés par des humains.
- Transparence : toute utilisation de l'IA pour la création de contenu est divulguée (par exemple, note de bas de page dans les publications).
- Exactitude et vérification des faits : les résultats de l'IA sont vérifiés par rapport aux sources primaires ; aucune diffusion d'informations non vérifiées.
- Éthique et équité : prévention des biais, de la discrimination ou de la désinformation ; respect de la LPD (loi sur la protection des données).
- Durabilité : utilisation limitée afin d'éviter toute dépendance ; révision annuelle des lignes directrices.
- Responsabilité personnelle : les auteur·e·s s'engagent à respecter la clause de déclaration d'originalité figurant dans l'ébauche d'idée/note de concept.

Utilisation dans les blogs et les publications

L'IA peut être utilisée pour la génération de texte, la recherche et la révision.

- Pouvons-nous utiliser l'IA ? Oui, par exemple pour des brouillons, des résumés ou la recherche d'idées (outils : ChatGPT, Claude, Gemini, Perplexity). Toujours indiquer : « Ce texte a été rédigé avec l'aide de l'IA et vérifié par [nom]. »
- Références à l'avenir et prévisions : dans le cas de scénarios spéculatifs (par exemple « L'IA dans la politique étrangère en 2030 »), les identifier clairement comme des hypothèses ; si l'IA présente quelque chose d'existant comme « nouveau », supprimer ou corriger – vérification des faits obligatoire.
- Exemple de note de bas de page : « Ébauche réalisée avec l'aide de GPT-4o ; version finale révisée et vérifiée par [nom]. »

Aspect	Autorisé	Non autorisé	Obligation de divulgation
Ébauche de texte	Oui		Toujours en note de bas de page
Recherche	Oui, pour générer des sources	Accepter comme seule source	« Recherche assistée par l'IA »
Traduction	Oui, avec vérification	Sans correction humaine	Facultatif
Scénarios d'avenir	Oui, en tant qu'hypothèse	Présenter comme un fait	« Scénario basé sur... »

Images et visuels générés par l'IA

Les images générées par l'IA (par exemple avec Midjourney, DALL-E) sont autorisées à des fins illustratives, mais pas pour des représentations factuelles.

- Règles : uniquement en cas de nécessité (par exemple, scénarios futuristes dans les blogs) ; les illustrations réalistes sont interdites afin d'éviter la désinformation.
- Mention obligatoire : toujours ajouter un filigrane ou une légende : « Image générée par l'IA via [outil] ». Pas d'utilisation dans les rapports officiels.
- Alternatives : privilégier les photos d'archives libres de droits ou vos propres graphiques ; recourir à l'IA uniquement si cela est indispensable.

- Exemple : dans un blog sur « L'IA dans la diplomatie » : une image abstraite de robots autour d'une table de négociation, avec la légende « Image générée par l'IA via DALL-E 3) ».

Gestion des risques et formation

- Évaluer les risques avant utilisation : contrôle des biais (par exemple, diversifier les requêtes), test de précision.
- Nouvelle entrée dans le guide pratique sur la plateforme des membres du foraus.
- Formation : atelier annuel sur les outils d'IA et l'éthique (inspiré de Digital Responsibility CH).
- Sanctions : une utilisation non déclarée de l'intelligence artificielle peut entraîner des révisions supplémentaires au cours du processus de traitement, voire le refus de la publication.

Mise en œuvre et mise à jour

Ces lignes directrices entrent en vigueur immédiatement ; l'équipe Engagement du bureau du foraus est chargé d'en informer les auteur·e·s. La mise en œuvre relève de la responsabilité des auteur·e·s et des relecteur·rice·s. Une évaluation annuelle sera réalisée sur la base des retours d'expérience et des nouvelles réglementations (par exemple, la loi européenne sur l'IA, les mises à jour au niveau fédéral). Pour toute modification : engagement@foraus.ch. Des lignes directrices sur l'IA mises à jour chaque année renforcent notre statut de think tank pertinent et digne de confiance.

Ressources pour les auteur·e·s

- [Recommendations on the Use of Generative Artificial Intelligence at UZH | University of Zurich](#)
- [AlgorithmWatch's guidelines to use generative AI responsibly](#)
- [Leitlinien zum Umgang mit Künstlicher Intelligenz in Kommunikation und Marketing an der Universität Bern](#)
- [EU AI Act](#)

EN: Guidelines on the use of AI at foraus

foraus permits the use of Artificial Intelligence (AI) tools in content production (blogs, publications, etc.) to improve efficiency, subject to strict conditions designed to safeguard foraus's credibility, transparency and independence as a foreign policy think tank. These guidelines are based on federal guidelines (transparency, non-discrimination, accountability),

EU ethical guidelines and foraus policy documents relating to AI governance. They apply to all staff and volunteers; responsibility lies with the lead author.

General principles

The use of AI is permitted where it improves quality and minimises risks. Fundamental principles:

- Human oversight: AI provides support but does not replace expertise; final content is always verified and approved by humans.
- Transparency: any use of AI for content creation is disclosed (e.g. footnote in publications).
- Accuracy and fact-checking: AI outputs are verified against primary sources; no dissemination of unverified information.
- Ethics and fairness: prevention of bias, discrimination or misinformation; compliance with the FADP (Data Protection Act).
- Sustainability: limited use to avoid dependency; annual review of the guidelines.
- Personal responsibility: authors undertake to comply with the originality declaration clause included in the draft idea/concept note.

Use in blogs and publications

AI can be used for text generation, research and proofreading.

- Can we use AI? Yes, for example for drafts, summaries or brainstorming (tools: ChatGPT, Claude, Gemini, Perplexity). Always state: ‘This text was written with the help of AI and checked by [name].’
- References to the future and predictions: in the case of speculative scenarios (e.g. ‘AI in foreign policy in 2030’), clearly identify them as hypotheses; if AI presents something that already exists as ‘new’, delete or correct it – fact-checking is mandatory.
- Example of a footnote: “Draft produced with the help of GPT-4; final version revised and verified by [name].”

Aspect	Permitted	Not permitted	Disclosure required
Text draft	Yes		Always in a footnote
Research	Yes, to generate sources	Accept as sole source	‘AI-assisted research’

Translation	Yes, yes with verification	Without human correction	Optional
Future scenarios	Yes, as a hypothesis	Presented as a fact	'Scenario based on...'

AI-generated images and visuals

Images generated by AI (e.g. using Midjourney, DALL-E) are permitted for illustrative purposes, but not for factual representations.

- Rules: only where necessary (e.g. futuristic scenarios in blogs); realistic illustrations are prohibited to prevent misinformation.
- Mandatory attribution: always add a watermark or caption: 'Image generated by AI via [tool]'. Do not use in official reports.
- Alternatives: prioritise royalty-free stock photos or your own graphics; use AI only if absolutely necessary.
- Example: in a blog post on 'AI in diplomacy': an abstract image of robots around a negotiating table, with the caption 'Image generated by AI via DALL-E 3'.

Risk management and training

- Assess risks before use: check for bias (e.g. by diversifying prompts), accuracy testing.
- New entry in the practical guide on the foraus members' platform.
- Training: annual workshop on AI tools and ethics (inspired by Digital Responsibility CH).
- Sanctions: undeclared use of artificial intelligence may result in additional reviews during the processing stage, or even refusal of publication.

Implementation and updating

These guidelines come into effect immediately; the GS Engagement department is responsible for informing authors of them. Authors and reviewers are responsible for ensuring compliance. An annual review will be carried out based on feedback and new regulations (e.g. the EU AI Act, updates at federal level). For any changes: engagement@foraus.ch. AI guidelines updated annually reinforce our role as a trusted Think tank.

Ressources pour les auteur·e·s

- [Recommendations on the Use of Generative Artificial Intelligence at UZH | University of Zurich](#)
- [AlgorithmWatch's guidelines to use generative AI responsibly](#)
- [Leitlinien zum Umgang mit Künstlicher Intelligenz in Kommunikation und Marketing an der Universität Bern](#)
- [EU AI Act](#)