



14/4/2026

AI Guidelines Policy for the Silat Wassel Platform

This policy outlines the ethical and operational framework for the use of artificial intelligence (AI) on the Silat Wassel platform. We are committed to using AI responsibly to foster communication while upholding the values of human dignity, trusted resources, and social equity. This policy is a living document and is continually updated to reflect technological advancements and emerging regulatory landscapes.

1. Our Definition and Approach to AI

We view AI not merely as a tool but as an interdisciplinary research field that replicates human intelligence via computational technology. Our approach is guided by:

Human-Centered Design & Hope-Based Communication

We prioritize a human and hope-based approach to all digital interactions. AI is utilized to bridge gaps in communication, ensuring that technology serves the dignity and aspirations of our audience rather than replacing the human element.

Critical Awareness of Socio-Political Context

We recognize that AI is not a neutral tool; it is deeply embedded in socio-political contexts. Because AI is trained on data that often represents dominant versions of the world, it can reproduce existing social biases. We remain critically aware of these inherent limits and the potential for harm in data representation.

Response-Ability & Human Oversight

We are committed to creating "ecosystems of response-ability." This means that we, as humans, hold ultimate responsibility for the outcomes, regulations, and effects of AI interactions. AI serves as an assistant for decision-making, but it does not carry final authority.

Operational Workflow: From Generation to Approval

To ensure consistency across all teams, the following workflow is mandatory for all AI-assisted tasks:

- 1. Drafting/Generation: Utilize the "Context + Instruction" model to prompt the AI.
- 2. Internal Review: Initial check by the creator for tone, mission alignment, and obvious errors.
- 3. Verification: Cross-referencing outputs with the "Trusted Resource Mapping" (see Section 4).
- 4. Final Approval: A designated team lead must review and approve the output before it is published or implemented.

2. Risk-Based Categorization and Oversight

We distinguish between different types of AI use to ensure that higher-risk applications receive appropriate scrutiny:

Risk Level	Definition	Examples	Level of Oversight
Low Risk	Administrative tasks with no public-facing impact.	Transcription, grammar checking, brainstorming internal meeting structures.	Individual user review for logic and clarity.

Medium Risk	Tasks that inform or assist in content creation.	Drafting social media posts, summarizing internal reports, and initial research.	Peer review; content must be verified against primary sources.
High Risk	Near-final outputs for public consumption or decision-making.	Generating policy recommendations, public-facing articles, or direct audience interaction.	Mandatory review by the AI expert and senior leadership sign-off.

3. Data Integrity and Bias Mitigation

Data is a symbolic abstraction of reality. To prevent the reproduction of biases, Silat Wassel implements the following safeguards:

- **Bias Identification:** We actively monitor for biases in allocation (e.g., gender-based selection) and service. Teams must explicitly prompt the AI to draw from diverse perspectives and look for gaps in the representation of subpopulations.
- **Representation:** We strive to ensure "human" is not defined by a narrow demographic.
- **Dignity:** We strictly prohibit the generation of content that makes offensive generalizations or dehumanizing comparisons.
- **Flagging Mechanism:** If a team identifies a persistent bias in a tool's output, they must escalate the issue to the AI expert to adjust the prompting strategy or tool selection.
 - ***Bias and Representation Check:***
 - Are different groups or perspectives fairly represented?
 - Is any group portrayed in a stereotypical or narrow way?
 - Has the AI defaulted to a dominant or Western-centric perspective?

Have we prompted or validated against diverse sources where relevant?

4. Fact-Checking & Trusted Resources

To combat the risks of AI "hallucinations," Silat Wassel implements a strict verification layer:

- **Manual Verification:** All AI-generated content intended for public consumption must be reviewed by the team to ensure accuracy.
- **Trusted Resource Mapping:** Our AI systems are directed to prioritize information from civil society manifestos, organized professional networks, and local partner data.
- **Step-by-Step Verification Workflow:**
 1. Identify key claims, statistics, or citations in the AI output.
 2. Verify each claim against a "Human-Verified" source or a trusted organizational database.
 3. If a claim cannot be verified through a trusted source, it must be **removed**.

5. The Silat Wassel AI Expert

We have developed a specialized AI expert to support both our team and our audience. This expert is trained to:

- **Address Inquiries:** Take any questions or suggestions to help improve platform operations.
- **Safe Interaction:** Operates under a "Context + Instruction" model to provide specific, high-quality guidance.
- **Consultation:** Teams must consult the AI expert for all "high-risk" tasks or when troubleshooting complex ethical dilemmas.
- **Privacy First:** Users are reminded never to share sensitive personal information with this or any AI tool.

You can reach out to our AI expert contact via:

Email: media@silatwassel.com or WhatsApp Phone Number: +20 10 405 982 38

6. Generative AI and Prompting Strategy

When using Generative AI (Gen AI), we prioritize the following:

- **Specificity over Randomness:** Using detailed prompting strategies to ensure outputs align with our mission.
- **Plausibility Checks:** Recognizing that Large Language Models (LLMs) predict patterns to generate "plausible" responses, which must always be fact-checked by our human team.

7. Accountability and Governance

Silat Wassel adheres to the principles of the Civil Society Manifesto on AI.

- **Decision Support:** AI should assist in making decisions, not replace human judgment.
- **Democratization:** We involve our diverse audience in the conversation regarding how these tools are used.
- **Living Document Governance:** * **Review Schedule:** This policy is reviewed and updated every six months.
 - **Responsibility:** The Silat Wassel administrative team and the AI expert are responsible for maintaining and updating these guidelines.
 - **Feedback Loop:** Teams are required to submit monthly feedback on AI tool performance to support continuous improvement.

النسخة العربية:

سياسة الذكاء الاصطناعي لمنصة "صلة واصل"

تحدد هذه السياسة الإطار الأخلاقي والتشغيلي لاستخدام الذكاء الاصطناعي في منصة صلة واصل. نحن ملتزمون باستخدام هذه التقنية بمسؤولية لتعزيز التواصل مع الحفاظ على قيم الكرامة الإنسانية، والمصادر الموثوقة، والعدالة الاجتماعية. تُعد هذه السياسة وثيقة حية ويتم تحديثها باستمرار.

١. تعريفنا ونهجنا تجاه الذكاء الاصطناعي

لا ننظر إلى الذكاء الاصطناعي كأداة فحسب، بل كمجال بحثي يحاول محاكاة الذكاء البشري. ويسترشد نهجنا بالآتي:

التصميم المتمحور حول الإنسان والتواصل القائم على الأمل
نمنح الأولوية للنهج البشري لضمان أن التكنولوجيا تخدم تطلعات جمهورنا ولا تحل محل العنصر البشري.

سير العمل العملي: من الإنشاء إلى الاعتماد

لضمان الاتساق، تتبع جميع الفرق سير العمل التالي:

١. الصياغة: استخدام نموذج "السياق + التعليمات".

٢. المراجعة الداخلية: التحقق الأولي من قبل منشئ المحتوى.

٣. التحقق: مقارنة المخرجات مع "خرائط المصادر الموثوقة".

٤. الاعتماد النهائي: مراجعة واعتماد من قبل قائد الفريق قبل النشر.

٢. تصنيف المخاطر والإشراف

- مخاطر منخفضة: (مثل التنسيق أو المسودات الداخلية) - مراجعة فردية.
- مخاطر متوسطة: (مثل مسودات وسائل التواصل أو تلخيص التقارير) - مراجعة الزملاء والتحقق من المصادر.
- مخاطر عالية: (مثل المحتوى الموجه للجمهور أو التوصيات السياسية) - مراجعة إلزامية من خبير الذكاء الاصطناعي.

٣. نزاهة البيانات والحد من التحيز

البيانات هي تجريد للواقع، ولذا نتبع الضمانات التالية:

- تحديد التحيز: مراقبة التحيزات بنشاط ومطالبة الذكاء الاصطناعي بتبني وجهات نظر متنوعة.
- التمثيل العادل: ضمان عدم تهميش أي فئات فرعية.

- الكرامة: منع إنتاج أي محتوى يتضمن تعميمات مسبقة.
- آلية الإبلاغ: يجب إبلاغ خبير الذكاء الاصطناعي عند رصد أي تحيز مستمر لتعديل استراتيجيات الأوامر (Prompts).

٤. تدقيق الحقائق والمصادر الموثوقة

لمكافحة مخاطر "الهوسة"، نتبع طبقة تحقق صارمة:

- المراجعة البشرية: مراجعة كل ما يوجه للجمهور.
- رسم خرائط المصادر الموثوقة: إعطاء الأولوية لموثيق المجتمع المدني والبيانات المحلية الموثوقة.
- خطوات التحقق: استخراج الادعاءات، التحقق منها من مصدرين مستقلين، وحذف أي معلومة لا يمكن إثباتها.

٥. خبير "صلة واصل" للذكاء الاصطناعي

- دعم العمليات: متاح للإجابة على التساؤلات وتحسين العمليات.
- الاستشارة: استشارة الخبير إلزامية في المهام عالية المخاطر.
- الخصوصية أولاً: يُمنع مشاركة أي معلومات شخصية حساسة مع أي أداة ذكاء اصطناعي.

للتواصل مع خبير الذكاء الاصطناعي في "صلة واصل":

البريد الإلكتروني: media@silatwassel.com أو

رقم هاتف الواتساب: +20 10 40598238

٦. المساءلة والحوكمة

- دعم القرار: الذكاء الاصطناعي يساعد في اتخاذ القرار ولا يحل محل التفكير البشري.
- حوكمة الوثيقة: يتم مراجعة هذه السياسة كل ستة أشهر من قبل الإدارة والخبير، مع اعتماد آلية للتغذية الراجعة الشهرية من الفرق الميدانية.

للمزيد من المعلومات أو لاقتراح تحديثات على هذه السياسة، يرجى التواصل مع الفريق الإداري لـ "صلة واصل" عبر البريد الإلكتروني: media@silatwassel.com