AI-ASSISTED SOFTWARE DEVELOPMENT POLICY

CHÍNH SÁCH PHÁT TRIỂN PHẦN MỀM CÓ HỖ TRỢ CỦA AI

Human-in-the-loop • Least Privilege • Traceability • Accountability

Tổ chức / Organization	[Tên tổ chức]
Chủ sở hữu / Document Owner	[Head of Engineering]
Phê duyệt / Approved by	Engineering Governance Council
Ngày hiệu lực / Effective Date	[DD/MM/YYYY]
Phiên bản / Version	1.0
Chu kỳ rà soát / Review Cycle	Hằng năm, hoặc ngay khi có công cụ/rủi ro AI mới đáng kể xuất hiện
Tài liệu cấp trên / Parent Documents	Engineering Constitution v1.0 (Tầng 1); Software Development Governance Policy v1.0 (Tầng 2)
Phân loại / Classification	INTERNAL

SAMPLE DOCUMENT

A. Kiểm soát tài liệu

DOCUMENT CONTROL

Phiên bản	Ngày	Người soạn	Mô tả thay đổi
1.0	[DD/MM/YYYY]	[Tên / Vai trò]	Ban hành lần đầu

B. Mục lục nội dung

CONTENTS

- 1. Mục đích
- 2. Phạm vi áp dụng
- 3. Liên kết với Engineering Constitution
- 4. Định nghĩa
- 5. Yêu cầu bắt buộc
- 6. Vai trò và trách nhiệm
- 7. Đo lường tuân thủ
- 8. Cơ chế ngoại lệ
- 9. Không tuân thủ
- 10. Chủ sở hữu tài liệu và chu kỳ rà soát
- 11. Tài liệu liên quan
- 12. Phê duyệt

1. Mục đích

PURPOSE

1. Cho phép tổ chức khai thác lợi ích năng suất của công cụ AI hỗ trợ lập trình (AI code assistant, AI coding agent) trong khi vẫn kiểm soát được rủi ro về bảo mật, chuỗi cung ứng, bản quyền, chất lượng và trách nhiệm giải trình.
2. Bảo đảm việc dùng AI trong phát triển phần mềm không làm suy yếu các kiểm soát đã có trong Software Development Governance Policy (SDGP) — đặc biệt là peer review và phân tách trách nhiệm.
3. Thiết lập ranh giới rõ ràng giữa việc dùng AI như một công cụ hỗ trợ con người và việc trao quyền tự chủ cho AI hành động thay con người trong hệ thống dùng chung hoặc production.

Chính sách này bổ sung cho SDGP, không thay thế — mọi yêu cầu của SDGP (SDLC, peer review, quản lý sự cố...) vẫn áp dụng đầy đủ khi công việc có sử dụng AI hỗ trợ.

2. Phạm vi áp dụng

SCOPE

- Mọi công cụ AI tạo sinh dùng trong hoạt động phát triển phần mềm: trợ lý gợi ý mã nguồn tích hợp IDE, chatbot lập trình, AI coding agent có khả năng tự thực hiện nhiều bước (đọc mã, sửa mã, chạy lệnh, mở pull request), và công cụ AI dùng để review hoặc tạo tài liệu kỹ thuật.
- Áp dụng cho toàn bộ vòng đời phần mềm nêu tại SDGP, từ viết mã, kiểm thử, review, đến vận hành.
- Áp dụng cho nhân sự nội bộ, nhà thầu và đối tác thuê ngoài khi họ phát triển phần mềm cho tổ chức.
- Không áp dụng cho việc dùng AI ngoài phạm vi phát triển phần mềm (ví dụ AI hỗ trợ marketing, chăm sóc khách hàng) — các hoạt động đó thuộc chính sách sử dụng AI khác của tổ chức, nếu có.

3. Liên kết với Engineering Constitution

ALIGNMENT WITH THE CONSTITUTION

Yêu cầu trong Policy	Nguyên tắc Constitution tương ứng
5.1 Quản lý công cụ AI (danh mục & pilot có kiểm soát)	Điều 2 — quản trị chuỗi cung ứng phần mềm/dịch vụ bên ngoài
5.2 Bảo vệ dữ liệu và giới hạn xử lý nội dung không tin cậy	Điều 2 — Security and Privacy by Design
5.3 Review của con người là bắt buộc	Điều 3, Điều 4 — Quality / Segregation of Duties
5.4 Minh bạch và khả năng truy vết	Điều 8 — Evidence over Opinion
5.5 Chuỗi cung ứng, bản quyền và mã nguồn do AI tạo	Điều 2, Điều 9 — Security by Design / Reuse có trách nhiệm
5.6 Trách nhiệm giải trình	Điều 4 — Segregation of Duties

5.7 Giới hạn quyền tự chủ của AI agent	Điều 4, Điều 7 — Segregation of Duties / Small, Reversible Changes
5.8 Hạn chế sử dụng cho tác vụ nhạy cảm	Điều 2 — least privilege, default deny
5.9 Đào tạo và nâng cao nhận thức	Điều 10 — Continuous Learning
5.10 Giám sát và ghi log	Điều 6 — Observable and Operable by Default

4. Định nghĩa

DEFINITIONS

- **AI Code Assistant:** công cụ AI gợi ý hoặc hoàn thiện đoạn mã theo yêu cầu tức thời của lập trình viên, hoạt động trong một phiên tương tác do con người khởi xướng và kiểm soát từng bước (ví dụ: gợi ý mã tích hợp IDE, chatbot lập trình).
- **AI Coding Agent (Agentic AI):** hệ thống AI có khả năng tự lập kế hoạch và thực hiện một chuỗi hành động nhiều bước với mức độ giám sát liên tục thấp hơn (đọc/sửa nhiều file, chạy lệnh, gọi công cụ bên ngoài, mở pull request) trước khi con người can thiệp lại.
- **Công cụ AI được phê duyệt (Approved AI Tool):** công cụ đã qua đánh giá rủi ro dữ liệu/bảo mật và được đưa vào danh mục Tier 1 theo Điều 5.1.
- **Rò rỉ dữ liệu qua AI (AI Data Leakage):** việc dữ liệu nội bộ, dữ liệu cá nhân hoặc bí mật kinh doanh bị đưa vào một công cụ AI theo cách có thể bị lưu trữ, ghi log hoặc dùng để huấn luyện mô hình ngoài tầm kiểm soát của tổ chức.
- **Ảo giác gói phần mềm / Slopsquatting (Package Hallucination):** hiện tượng AI đề xuất tên một thư viện/dependency không có thật; kẻ tấn công có thể chủ động đăng ký gói mang đúng tên đó kèm mã độc để chờ nạn nhân cài đặt.
- **Con người giám sát (Human-in-the-loop):** yêu cầu một con người có đủ thẩm quyền phải xem xét và chấp thuận trước khi một hành động do AI đề xuất có hiệu lực trên hệ thống dùng chung.

5. Yêu cầu bắt buộc

MANDATORY REQUIREMENTS

5.1 Quản lý công cụ AI: danh mục phê duyệt và pilot có kiểm soát

1. Công cụ AI dùng trực tiếp với mã nguồn, dữ liệu hoặc hệ thống chính thức của tổ chức (Tier 1) **PHẢI (MUST)** nằm trong danh mục đã được Engineering Governance Council hoặc bộ phận được ủy quyền phê duyệt, sau khi đã đánh giá tối thiểu: chính sách dữ liệu của nhà cung cấp (có dùng dữ liệu để huấn luyện mô hình hay không), nơi lưu trữ dữ liệu, và khả năng tắt tính năng lưu giữ dữ liệu (data retention opt-out) nếu có.
2. Để không cản trở việc thử nghiệm công nghệ mới, kỹ sư **ĐƯỢC PHÉP (MAY)** đăng ký thử nghiệm một công cụ AI mới (Tier 2 — Sandbox/Pilot) mà không cần chờ phê duyệt đầy đủ như Tier 1, với điều kiện: chỉ dùng trong môi trường cô lập, không có dữ liệu thật, không có credentials/API key thật, và không kết nối trực tiếp vào kho mã nguồn chính thức.
3. Nhân viên **KHÔNG ĐƯỢC PHÉP (MUST NOT)** dùng bất kỳ công cụ AI nào ngoài hai hình thức trên (Tier 1 đã duyệt, hoặc Tier 2 đã đăng ký trong sandbox) cho công việc thuộc phạm vi Policy này.

5.2 Bảo vệ dữ liệu và giới hạn xử lý nội dung không tin cậy

1. **KHÔNG ĐƯỢC PHÉP (MUST NOT)** đưa vào bất kỳ công cụ AI nào (kể cả Tier 1, trừ khi cấu hình riêng cho phép): credentials, khóa API, dữ liệu cá nhân của khách hàng, hoặc dữ liệu được phân loại Bảo mật/Tối mật theo Data Classification Standard.
2. Mã nguồn của hệ thống được phân loại nhạy cảm cao chỉ được dùng với công cụ AI đã xác nhận không lưu giữ/huấn luyện trên dữ liệu đầu vào (theo Điều 5.1.1).
3. Khi cấp cho AI coding agent quyền đọc nội dung không do tổ chức kiểm soát (issue/ticket từ người dùng ngoài, mã nguồn mở bên thứ ba chưa qua rà soát, tài liệu tải từ Internet), **PHẢI (MUST)** giới hạn quyền ghi/thực thi (write/execute) của agent trong phiên xử lý nội dung đó, nhằm giảm rủi ro chỉ thị ẩn giấu trong nội dung (indirect prompt injection) khiến agent thực hiện hành động ngoài ý muốn.

5.3 Review của con người là bắt buộc

1. Mọi mã nguồn hoặc cấu hình do AI tạo ra **PHẢI (MUST)** được một con người đủ thẩm quyền review trước khi sáp nhập vào luồng mã chính, theo đúng yêu cầu peer review tại SDGP Điều 5.4. Việc một công cụ AI khác đã "kiểm tra" một thay đổi không thay thế được yêu cầu này.
2. Người review **PHẢI (MUST)** kiểm tra riêng tính toàn vẹn của bộ kiểm thử và cấu hình CI/CD liên quan đến thay đổi — bảo đảm AI không tự ý xóa, vô hiệu hóa, làm yếu test case, hoặc viết test không có khả năng phát hiện lỗi (test rỗng) chỉ để vượt qua pipeline.

5.4 Minh bạch và khả năng truy vết

1. Pull request có phần đóng góp đáng kể từ AI coding agent (tự thực hiện nhiều bước không có giám sát liên tục) **NÊN (SHOULD)** được ghi chú rõ trong mô tả thay đổi.
2. Với hệ thống được phân loại rủi ro cao, tổ chức **CÓ THỂ (MAY)** yêu cầu ghi nhận bắt buộc công cụ AI đã dùng để hỗ trợ, phục vụ điều tra khi cần.

5.5 Chuỗi cung ứng, bản quyền và mã nguồn do AI tạo

1. Mọi thư viện/dependency mới do AI đề xuất **PHẢI (MUST)** được xác minh là có thật (tồn tại trên registry chính thức, có lịch sử phát hành và cộng đồng sử dụng hợp lý) trước khi cài đặt, nhằm phòng tránh rủi ro ảo giác gói phần mềm (slopsquatting) nêu tại Điều 4.
2. Dependency do AI đề xuất **PHẢI (MUST)** được quét bằng công cụ Software Composition Analysis (SCA) theo cùng yêu cầu tại SDGP Điều 5.10, không có ngoại lệ vì "do AI chọn".
3. Mã nguồn do AI tạo ra **PHẢI (MUST)** được xử lý rủi ro bản quyền/giấy phép tương tự mã nguồn mở lấy từ bên ngoài — không mặc định coi là "sạch" bản quyền chỉ vì do AI viết ra. Không được cố ý yêu cầu AI sao chép nguyên văn mã nguồn có bản quyền của bên thứ ba.

5.6 Trách nhiệm giải trình

1. Người đề xuất một thay đổi — kể cả khi phần lớn nội dung do AI tạo ra — là người chịu trách nhiệm về tính đúng đắn, bảo mật và chất lượng của thay đổi đó.
2. "AI tạo ra" hoặc "AI đề xuất" **KHÔNG ĐƯỢC PHÉP (MUST NOT)** được dùng làm lý do biện minh khi xảy ra sự cố hoặc lỗi chất lượng trong post-incident review theo SDGP Điều 5.13.

5.7 Giới hạn quyền tự chủ của AI coding agent

1. AI coding agent **KHÔNG ĐƯỢC PHÉP (MUST NOT)** tự merge pull request do chính nó tạo ra vào bất kỳ nhánh dùng chung nào. Mọi thay đổi do agent tạo ra phải dừng ở trạng thái chờ (draft/open PR) và tuân theo đúng quy trình review tại Điều 5.3 — không có ngoại lệ "nhánh phụ" hay môi trường non-production.
2. Quyền truy cập hệ thống, kho mã nguồn và dữ liệu của AI coding agent **PHẢI (MUST)** áp dụng nguyên tắc đặc quyền tối thiểu (least privilege), giới hạn theo phạm vi nhiệm vụ cụ thể và có thời hạn.
3. AI coding agent hoạt động tự động, ít giám sát **PHẢI (MUST)** được giới hạn tài nguyên (số bước thực hiện, hạn mức token/chi phí, thời gian chạy tối đa) để tránh rủi ro tiêu tốn tài nguyên không kiểm soát khi agent rơi vào vòng lặp xử lý lỗi.

5.8 Hạn chế sử dụng cho tác vụ nhạy cảm

1. **KHÔNG ĐƯỢC PHÉP (MUST NOT)** để AI tự đưa ra quyết định cuối cùng (không có con người xác nhận) đối với: cấp quyền truy cập, phê duyệt thay đổi production, hoặc các quyết định ảnh hưởng trực tiếp đến quyền lợi cá nhân của người dùng/nhân viên.
2. Việc dùng AI hỗ trợ đánh giá bảo mật (ví dụ rà soát lỗ hổng) **PHẢI (MUST)** có chuyên gia bảo mật xác nhận lại kết quả trước khi kết luận một hệ thống là an toàn.

5.9 Đào tạo và nâng cao nhận thức

Nhân sự dùng công cụ AI hỗ trợ lập trình **PHẢI (MUST)** hoàn thành hướng dẫn sử dụng AI an toàn (bao gồm nhận thức về prompt injection và ảo giác gói phần mềm), như một phần của chương trình đào tạo tại SDGP Điều 5.12.

5.10 Giám sát và ghi log

Việc dùng AI coding agent có quyền thực thi lệnh hoặc thay đổi hệ thống **NÊN (SHOULD)** được ghi log tương xứng với mức độ quyền hạn được cấp, phục vụ điều tra khi cần.

> **Không lặp lại nội dung của SDGP:** Các yêu cầu về SDLC, peer review kỹ thuật, quản lý sự cố, đo lường tuân thủ và cơ chế ngoại lệ đã được quy định đầy đủ tại SDGP. Policy này chỉ bổ sung phần đặc thù của AI; khi có

> xung đột hoặc trùng lặp diễn giải, bản quy định tại SDGP được ưu tiên áp dụng cho phần chung.

6. Vai trò và trách nhiệm

ROLES & RESPONSIBILITIES

- **Engineering Governance Council:** phê duyệt danh mục công cụ AI Tier 1; phê duyệt sửa đổi Policy này.
- **Security/Risk:** đánh giá rủi ro dữ liệu của từng công cụ AI trước khi đưa vào danh mục Tier 1; xét duyệt đăng ký Tier 2 (Sandbox/Pilot).
- **Engineering Leadership:** ban hành Standard/Guideline hướng dẫn dùng AI cụ thể theo từng công cụ được phê duyệt.
- **Engineering Teams:** tuân thủ Policy, hoàn thành đào tạo, chịu trách nhiệm về mọi thay đổi mình đề xuất dù có dùng AI hỗ trợ.

7. Đo lường tuân thủ

COMPLIANCE MEASUREMENT

- Tỷ lệ công cụ AI đang dùng thực tế nằm trong danh mục Tier 1 hoặc đăng ký Tier 2 hợp lệ.
- Tỷ lệ pull request có nguồn gốc từ AI coding agent được review đầy đủ theo Điều 5.3.
- Số lần phát hiện dependency không tồn tại (package hallucination) trước khi cài đặt.
- Số sự cố rò rỉ dữ liệu qua công cụ AI được ghi nhận.
- Tỷ lệ nhân sự hoàn thành đào tạo dùng AI an toàn.

8. Cơ chế ngoại lệ

EXCEPTIONS

Áp dụng theo đúng cơ chế và giới hạn tại Engineering Constitution, Điều 8 và SDGP, Điều 8.

9. Không tuân thủ

NON-COMPLIANCE

Áp dụng theo đúng nguyên tắc tại SDGP, Điều 9: báo cáo và khắc phục cho vi phạm thông thường; chỉ xem xét kỷ luật lao động — qua HR/Pháp chế, theo Nội quy lao động — đối với hành vi cố ý (ví dụ: cố tình dùng công cụ AI ngoài danh mục với dữ liệu nhạy cảm, hoặc che giấu việc dùng AI khi được yêu cầu minh bạch tại Điều 5.4).

10. Chủ sở hữu tài liệu và chu kỳ rà soát

DOCUMENT OWNERSHIP & REVIEW CYCLE

Document Owner: [Head of Engineering]. Chu kỳ rà soát: hằng năm, hoặc ngay khi có công cụ/rủi ro AI mới đáng kể xuất hiện — lĩnh vực này thay đổi nhanh hơn nhiều tài liệu Tầng 2 khác nên chu kỳ rà soát cần linh hoạt hơn mức tối thiểu.

11. Tài liệu liên quan

RELATED DOCUMENTS

- Engineering Constitution v1.0, Software Development Governance Policy v1.0 (Tầng 1-2)
- Secure Coding Standard, Data Classification Standard, Data Protection Standard, Vulnerability Management Standard (Tầng 3)
- Danh mục công cụ AI được phê duyệt (Approved AI Tool List), AI Sandbox/Pilot Registration Procedure, AI Incident Response Procedure (Tầng 4/5)

AI là thực tập sinh không bao giờ mệt, không phải người ra quyết định cuối cùng;
mọi mã nguồn do AI tạo ra vẫn cần một con người hiểu và chịu trách nhiệm;
mọi dependency do AI đề xuất đều phải xác minh là có thật;
mọi quyền hạn cấp cho AI agent đều có giới hạn, có thời hạn và có thể thu hồi;
và không lời giải thích nào bắt đầu bằng "vì AI bảo vậy".

12. Phê duyệt

APPROVAL

Vai trò	Họ tên	Chữ ký	Ngày
Chair, Engineering Governance Council			
Chief Technology Officer			
Head of Engineering (Document Owner)			
Head of Technology Risk / Security			