Chính sách Quản trị AI — Layer 1 (Foundation)

GENERATED từ 10-rules — không sửa tay

Chính sách toàn công ty — áp cho mọi nhân viên, mọi phòng ban, mọi lúc. Nguyên tắc: mức kiểm soát tăng theo mức rủi ro. Dưới đây là các quy định tối thiểu (lằn ranh đỏ) áp cho tầng Foundation (L1).

Danh tính & công cụ được duyệt

AUTH-001 — Chỉ dùng công cụ AI được công ty duyệt

tiers: L1, L2, L3, L4

Chỉ công cụ AI **có hợp đồng công ty và input không bị dùng để train** mới được dùng cho việc công ty. **Tài khoản AI cá nhân bị cấm tuyệt đối** cho mọi dữ liệu công ty ("Shadow AI", L1 §8).

Danh sách hiện hành (`30-registry/ai-tools.yml`):

- Google Workspace Gemini (company)
- Claude (Team / Enterprise)

AUTH-002 — Mọi công cụ phải đăng nhập bằng tài khoản @vitalify.asia

tiers: L1, L2, L3, L4

Mọi công cụ dùng cùng AI — công cụ AI (`AUTH-001`) và công cụ làm việc / connector — phải xác thực bằng tài khoản `@vitalify.asia`; cấm tài khoản cá nhân — vì IT không thu hồi được quyền khi nhân viên nghỉ, không enforce được quyền sở hữu dữ liệu, và không có audit log để điều tra sự cố.

Danh sách hiện hành (`30-registry/connectors-mcp.yml`):

- Figma
- Google Docs/Sheets/Slides/Drive
- Draw.io

AUTH-003 — Xác nhận công cụ company-managed và dữ liệu được phép trước khi nhập

tiers: L1, L2, L3, L4

Trước khi nhập bất kỳ thông tin **không công khai** nào, phải xác nhận **cả hai**: công cụ do công ty quản lý (có hợp đồng, input loại khỏi training) **và** dữ liệu được phép theo chính sách. Thiếu một trong hai → dừng và hỏi PMO. *Trả tiền không đồng nghĩa an toàn.*

AUTH-004 — Thêm công cụ AI mới phải qua quy trình duyệt

tiers: L1, L2, L3, L4

Không tự ý dùng công cụ AI mới. Luồng duyệt: nhân viên đề xuất → quản lý trực tiếp → CoE tổng hợp → **Dev Director** quyết → CoE thông báo lại. Trước khi được duyệt, coi như **chưa được phép**.

Dữ liệu được / không được đưa vào AI

DATA-001 — Cấm tuyệt đối đưa dữ liệu khách hàng / secret production vào AI

tiers: L1, L2, L3, L4

Không bao giờ nhập: credential production/test (ID, mật khẩu, API key), dữ liệu cá nhân người dùng cuối / dữ liệu thô của khách (log test, lịch sử giao dịch), và thông tin tối mật nội bộ (tài chính chưa công bố, M&A). Đây là **lằn ranh đỏ số 1** của công ty. Ở L2 được chặn một phần qua `deny` của settings.json (Read/Write các mẫu file secret).

DATA-002 — Phân loại dữ liệu 3 mức

tiers: L1, L2, L3, L4

Mức	Có được nhập?
Tối mật (credential, PII người dùng cuối, M&A)	❌ Không — kể cả trên công cụ công ty (xem `DATA-001`)
Dữ liệu nội bộ (biên bản, đề xuất, spec)	⚠️ Có — **chỉ công cụ công ty**. Danh từ riêng như tên KH được nhập nếu là thông tin thường chia sẻ nội bộ; gặp ngoại lệ dự án → hỏi quản lý
Thông tin chung (public, tra cứu, soát lỗi văn bản không mật)	✅ Có — không hạn chế

DATA-003 — Ẩn danh hoá theo mức nhạy cảm (không ẩn danh toàn bộ)

tiers: L1, L2, L3, L4

COE chốt (#2): theo L1, nới luật ẩn danh hoá của L2. Không bắt buộc ẩn danh *mọi* tên KH. Cụ thể:

- Tên KH / danh từ riêng **được nhập như thực** nếu là thông tin thường chia sẻ nội bộ (vd qua email/chat).
- RFP / spec / đề xuất **được upload** sau khi đã gỡ secret, PII, IP nội bộ (xem `DATA-004`) — không cần thay bằng placeholder.
- Chỉ **bắt buộc ẩn danh** số liệu thương mại nhạy cảm (giá, rate, budget, margin) và dữ liệu thuộc `DATA-001`/`DATA-002` mức tối mật.

DATA-004 — Quy trình upload tài liệu (RFP / spec / design)

tiers: L1, L2, L3, L4

Trước khi upload tài liệu vào AI: **(1)** xác nhận hợp đồng cho phép dùng AI cho tài liệu đó — không chắc thì hỏi khách hoặc không dùng; **(2)** gỡ mật khẩu, API key, IP address nội bộ, dữ liệu cá nhân; **(3)** chỉ dùng công cụ đã duyệt. Trường hợp nhạy cảm review với PMO.

DATA-007 — Code có AI hỗ trợ phải sạch về IP

tiers: L1, L2, L3, L4

Code do AI sinh và đi vào sản phẩm bàn giao phải được **review và quét license/originality** (tránh tái tạo code GPL/AGPL làm "nhiễm" bản quyền), đạt cùng chuẩn chất lượng & bảo mật như code người viết. Điều khoản sở hữu IP của khách áp cho code AI y như mọi code khác.

DATA-008 — Không dùng generative AI cho brand asset thương mại

tiers: L1, L2, L3, L4

Không dùng AI sinh thiết kế thương hiệu cuối cùng giao khách hoặc làm nhận diện công ty (logo, mascot, trademark) — vì AI có thể sao chép tác phẩm được bảo hộ mà không phát hiện được, và không thể verify đã "sạch". AI được dùng cho cảm hứng/moodboard/nháp; bản cuối do người thiết kế làm. Khách yêu cầu thì escalate PMO.

DATA-011 — Chỉ gửi nội dung công việc cho công cụ AI đã duyệt

tiers: L1, L2, L3, L4

Không gửi source code, tài liệu, dữ liệu hay log tới bất kỳ dịch vụ bên ngoài nào khác ngoài công cụ AI đã duyệt mà bạn đang dùng — không pastebin, công cụ format/convert online, API AI/LLM khác, công cụ test webhook hay link chia sẻ file công khai. Khi nhiều công cụ AI cùng xử lý một nội dung, từng công cụ đều phải được duyệt (`AUTH-001`). Agent chỉ fetch từ domain đã duyệt (`RULES-006`).

Cấu hình công cụ AI an toàn

RULES-007 — AI không được tự thao tác trên hệ thống production, công ty hay khách hàng

tiers: L1, L2, L3, L4

AI không được tự hành động trên hệ thống của công ty, của khách hàng hay môi trường Production: không gửi email, gọi API hay vận hành hệ thống cloud khi chưa có người duyệt. AI không bao giờ được trực tiếp truy cập, vận hành, sửa đổi, deploy lên hay chạy lệnh nhắm vào Production — gồm script deploy và pipeline CI/CD, file cấu hình và biến môi trường, credential/secret/token/API key, hạ tầng, database, tài nguyên cloud và hệ thống vận hành. Kể cả khi script, credential hay cấu hình Production có sẵn trong source code/repo, AI chỉ được **đọc và đề xuất thay đổi**, không được chạy, áp dụng, thử hay dùng chúng với Production. Cụ thể, AI không được: chạy script deploy hoặc migration Production; chạy lệnh bằng credential Production; sửa hạ tầng, database, cấu hình hay secret Production; tắt hoặc vượt qua control bảo mật Production; kích hoạt job CI/CD nhắm vào Production. Mọi thao tác Production phải được người có thẩm quyền review, duyệt và **tự tay thực hiện**.

Ứng phó sự cố

RES-001 — Báo sự cố bảo mật ngay lập tức

tiers: L1, L2, L3, L4

Khi phát hiện sự cố (lộ dữ liệu, secret/credential bị đẩy lên GitHub, dữ liệu khách bị lộ, truy cập đáng ngờ…): **báo ngay** cho quản lý trực tiếp **và** CoE cùng lúc, tại kênh `[SECURITY] Incident Report`, ngay khi thấy — không chờ xác minh đủ, không tự sửa trước. Mô tả ngắn: chuyện gì, ở đâu (repo/tool/hệ thống), thấy khi nào.

RES-002 — Khi sự cố: dừng tác vụ, thu hồi quyền, truy vết phạm vi

tiers: L1, L2, L3, L4

Ngoài việc báo cáo, quy trình ứng phó phải gồm: **dừng** tác vụ/agent đang chạy, **thu hồi** access của agent/credential liên quan, và **truy vết** phạm vi ảnh hưởng về tới người/lệnh đã kích hoạt. L1 §11 hiện mới có "report cho manager + CoE" — **thiếu** bước stop/revoke/trace mà Kato yêu cầu bổ sung.