

Human-in-the-loop

en feature, ikke et kompromis

Af Velatir co-founder & CEO, Michael Sørensen

Agenda



Status Quo

Hvor mange AI-opsætninger fejler



Human-in-the-Loop og Compliance

Hvorfor det er nøglen til mere AI



Implementering og next steps

Hvordan gør man det i praksis?

Fra 2023 til 2024 steg danske virksomheders AI anvendelse med ~12 procentpoint. Fra 15% til 27%

kilde: eurostat

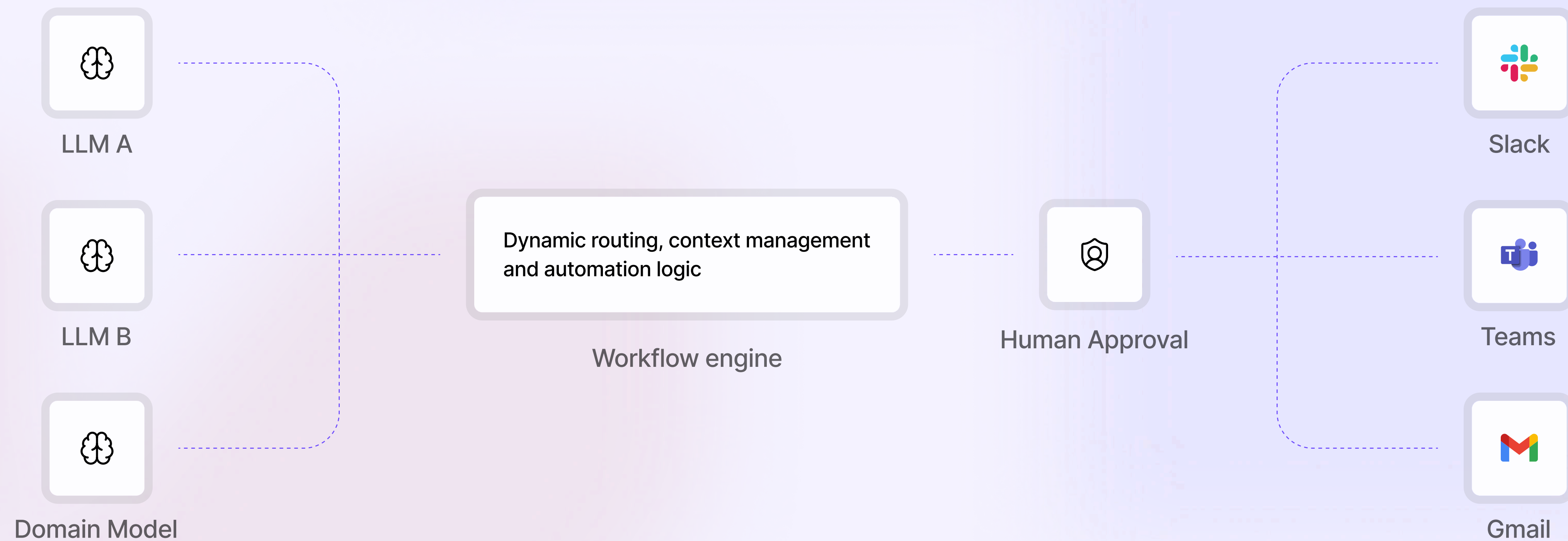
Størstedelen af de 27% er...

Siloede løsninger, Shadow AI og
Leverandør lock-in

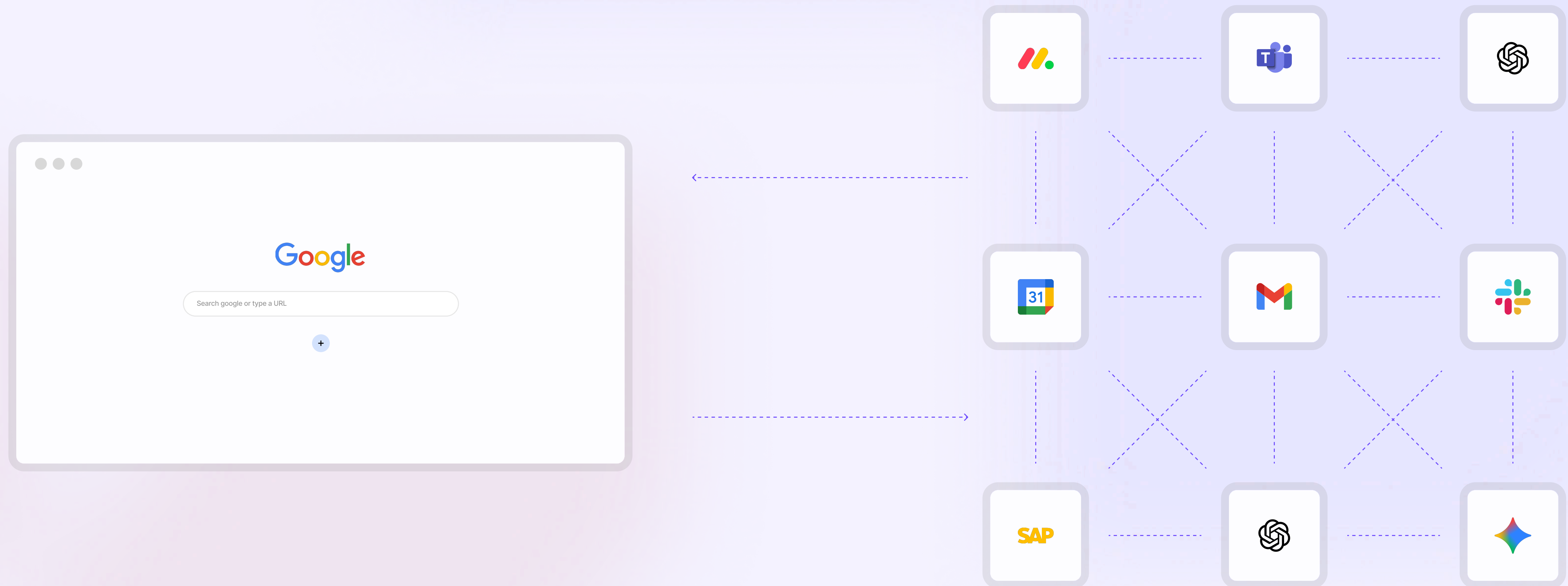


Hvor er vi på vej hen

Hybrid intelligens



Mindre context-switching, mere effektivitet, mere sikkerhed



Gevinster ved HITL og Guardrails

- ✓ Dybere integration (Sæt AI fri i organisationen)
- ✓ Størrere præcision
- ✓ Brugeraccept
- ✓ Mindre vedligeholdelse
- ✓ Compliance

Hvorfor Guardrails?

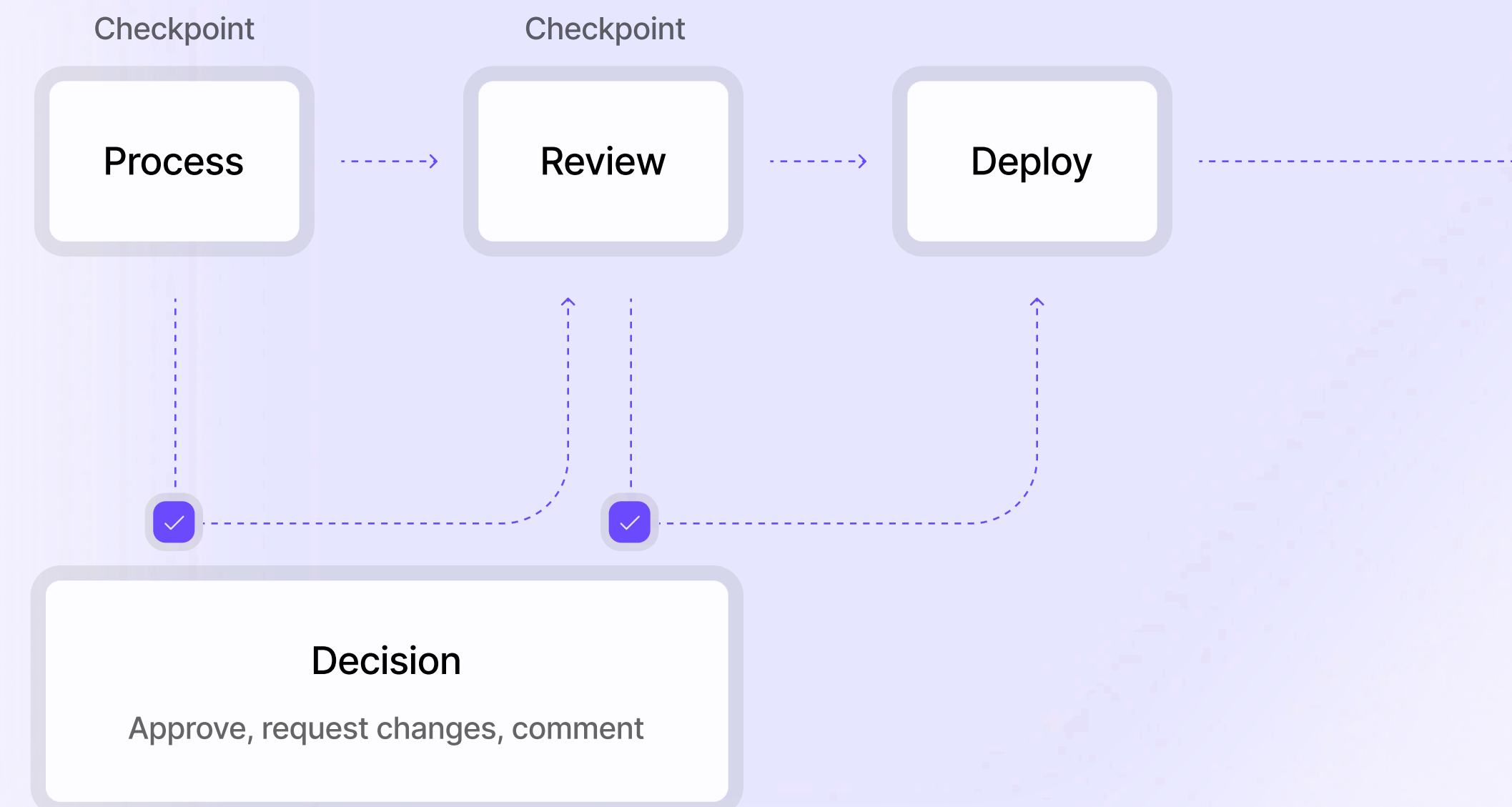
Compliance krav	Formål	Hvordan Human-in-the-Loop opfylder kravet	Praktiske eksempler
Transparency	Brugere og tilsyn skal kunne forstå og forklare AI-beslutninger.	HITL giver sporbare beslutninger: hvert menneskeligt input og modeloutput logges og kan forklares retrospektivt.	Approvers får kontekst og kan kommentere modelbeslutning → alle interaktioner gemmes.
Auditability	Organisationen skal kunne dokumentere og efterprøve AI-processer.	HITL-punkter danner naturlige audit-checkpoints med tidsstempler, godkendelser og modelversioner.	Audit-log viser: “model v1.2 foreslog X → godkendt af bruger Y kl. 14:32”.
Human oversight	Sikrer, at mennesker kan gribe ind, forhindre skade og sikre korrekt anvendelse.	HITL giver reel menneskelig kontrol og mulighed for override, især i high-risk-scenarier.	Eksempel: automatisk kreditvurdering → manuel godkendelse før endelig afgørelse.
Accountability	Der skal være tydelig ansvarsfordeling og beslutningshistorik.	HITL-sign-offs dokumenterer hvem der traf beslutningen, hvilket reducerer “black box”-problemet.	Godkendelsesflow med signatur / Slack-approve = ansvar knyttet til specifik bruger.
Data & Risk Management (NIS2/GDPR/DORA/AI Act)	Reducere risiko for datalæk, fejl og misbrug af automatiserede systemer.	Menneskelige checkpoints fanger fejl og misbrug tidligt og kan trigge incident-respons.	Hvis model output indeholder følsomme data → menneskelig reviewer afviser og rapporterer.

Designprincipper for guardrails og HITL i jeres workflows

Checkpoints på Input, modifikation, adgang og output.

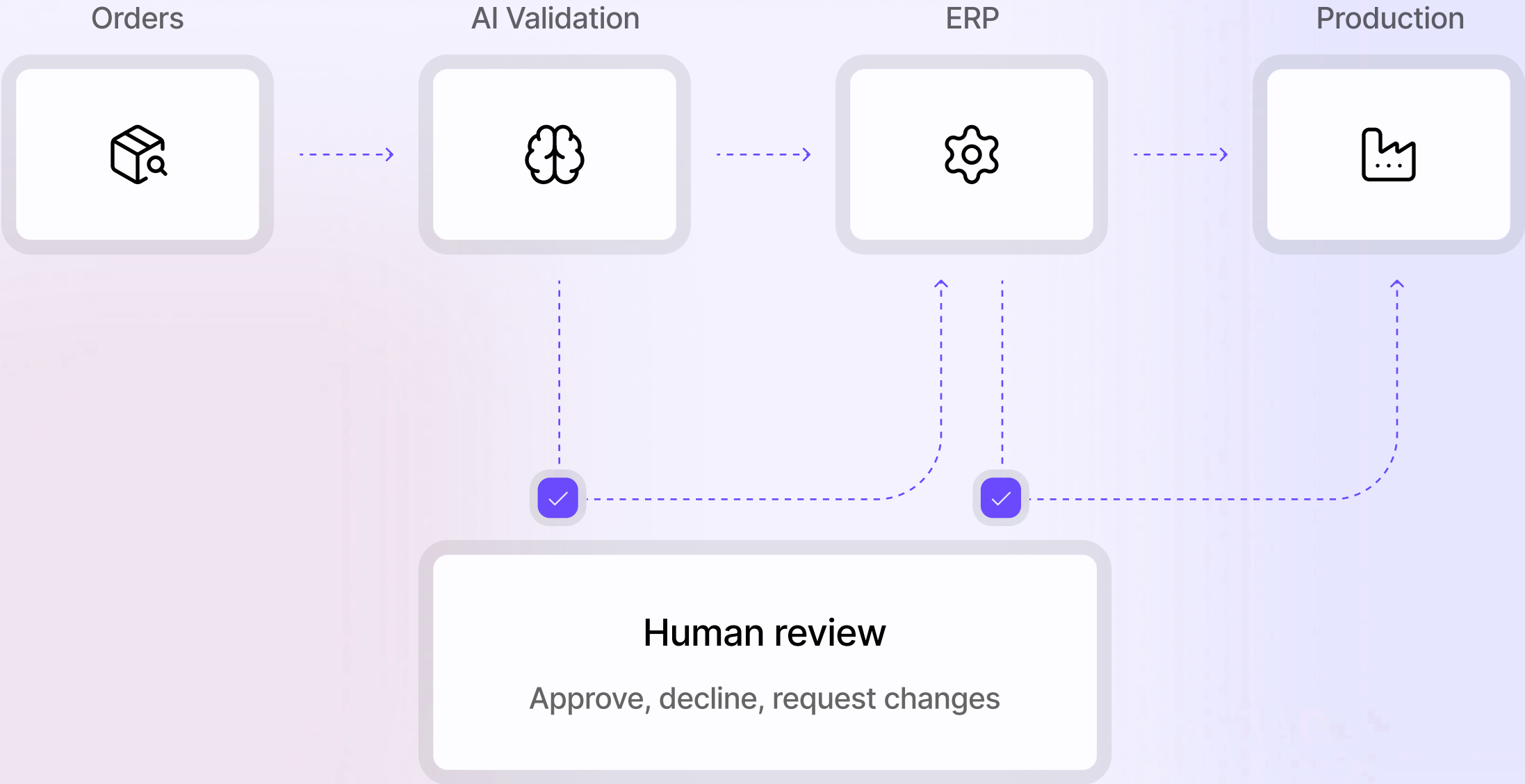
Log på alle checkpoints

Dynamisk eskalation - kontekstuel review (hvem, hvad, hvor, hvornår)



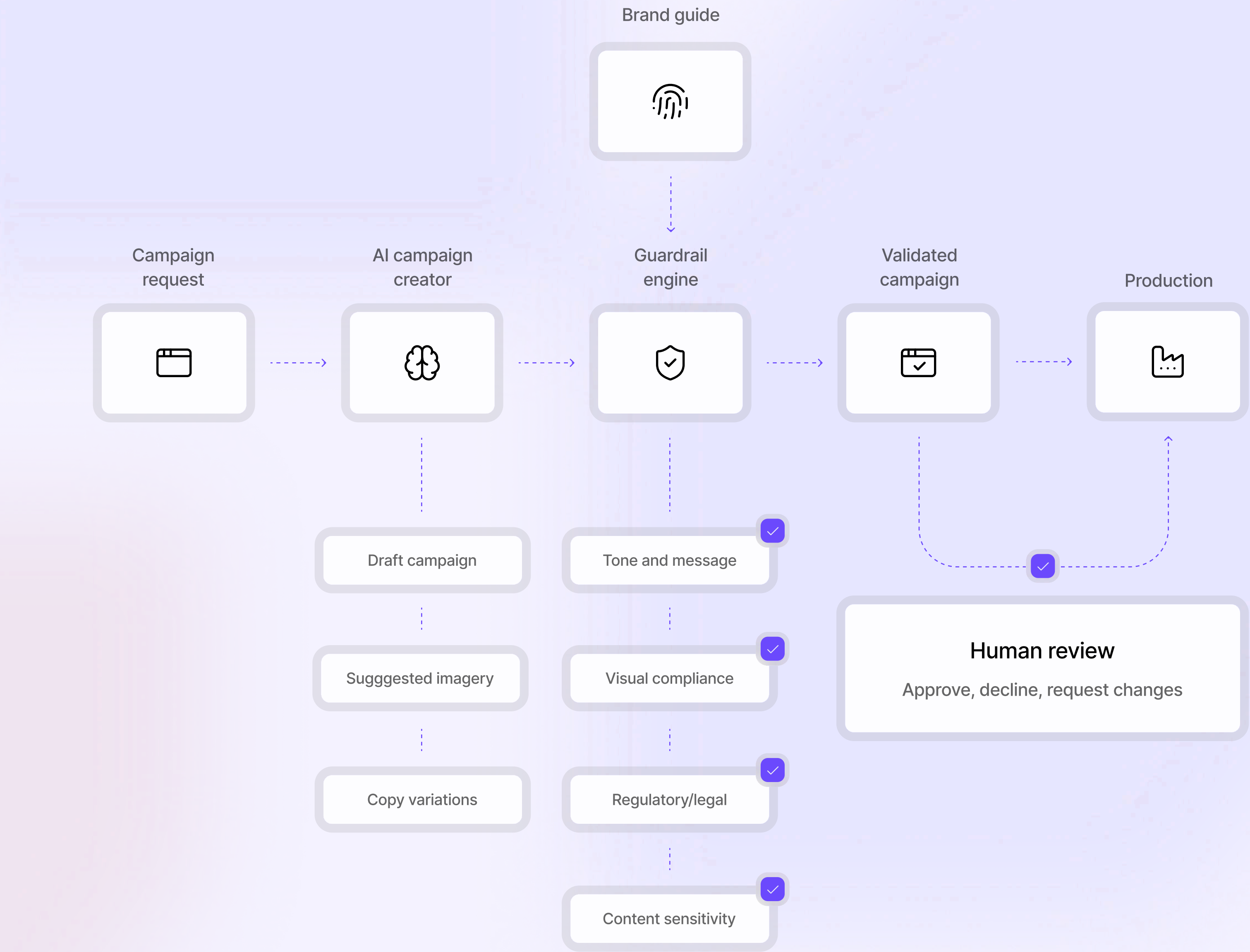
Case

Data Validering til ERP



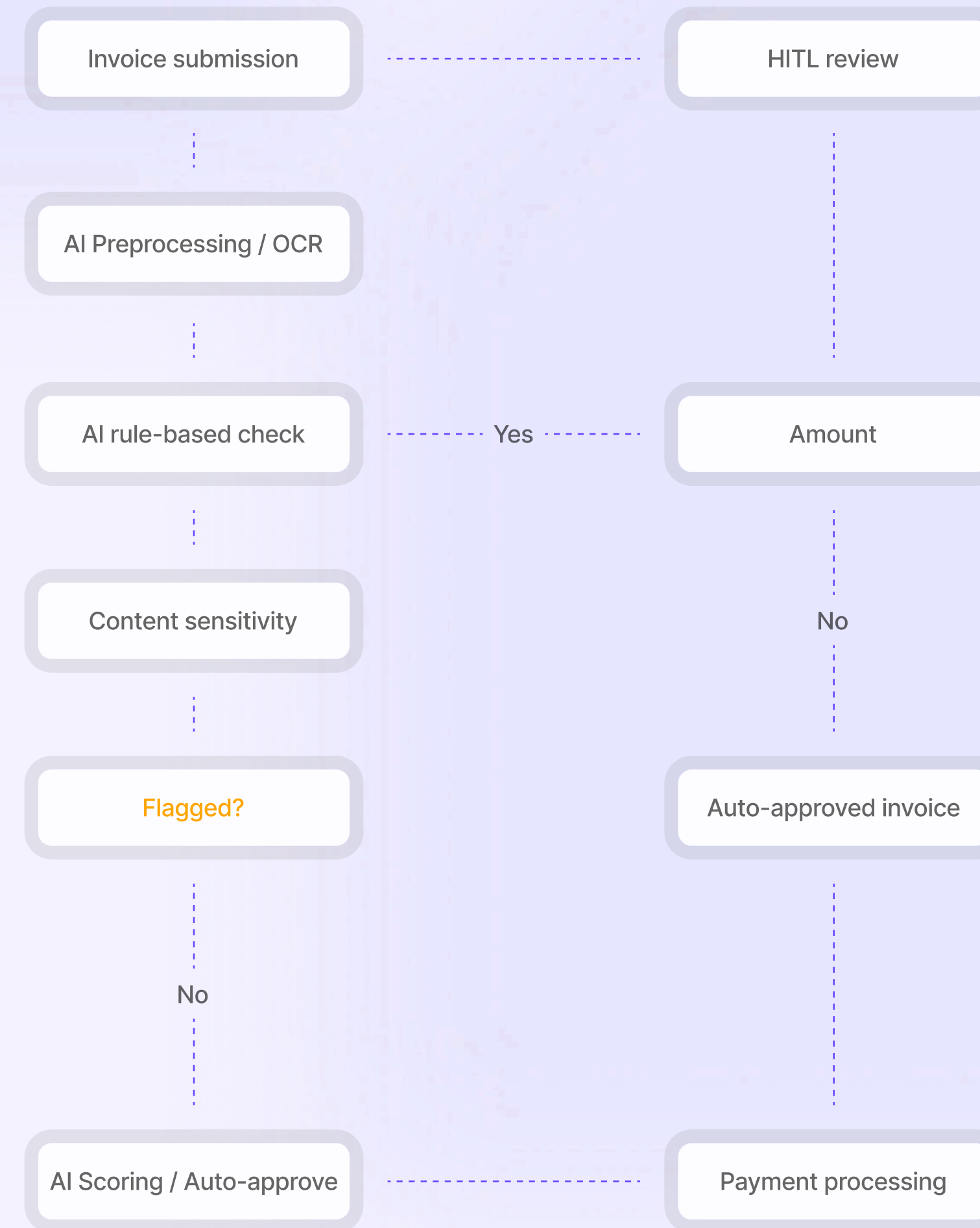
Case

Brand guide agent



Case

Fakturahåndtering



Opsummering

- ✓ HITL/Guardrails gør dig i stand til at sætte AI fri og minimere risiko.
- ✓ HITL/Guardrails gør dig compliance-ready.
- ✓ Guardrails og HITL giver dig data du kan iterere på omkring din AI og automationer.