『AI エージェントのローカル導入セキュリティチェックリスト』より

導入前に確認したい 10 の質問

それぞれ、確認する理由を添えています。

01 権限はモデルが資料を見る前に判定するか、それともモデル任せか

モデルに権限を判断させると、一言で回避されかねません。検索の後、生成の前にログイン身分で絞り込み、帰属が判定できない資料は既定で見せないようにします。

02 資料やツールの返り値に含まれる指示を、命令として実行してしまわないか

間接インジェクションはローカル小型モデル共通の弱点です。モデルの外で入力を除去し、資料はデータとして扱い、指示としては扱いません。

03 モデルが呼べるツールは何か。書き込みや機器制御のツールはあるか

ツールは登録し、等級を分けます。読み取りはモデルに任せてよく、書き込みや制御は別途の許可が必要で、できればモデルから見えないようにします。

04 回答の条文、数値、出典を一つずつ確かめられるか

結論はすべて元の資料に結び付けます。条文が見つからなければ「未収録」と書き、モデルに補わせません。数値はコードで計算し、モデルに暗算させません。

05 証拠が足りないとき、無理に答えるか、断るか、人に回すか

断れる、人に回せるシステムだけが、自分が知らないことを知っています。

06 人の審査は本当に人が承認したのか、既定で通るのか

人が確認したときだけ承認とし、裁定者はログイン身分で記録して、代理署名を認めません。

07 問題が起きたとき、誰が何に基づいて何をしたか説明できるか

手順ごとに記録し、改ざんの有無を検証できる台帳が必要です。

08 モデルが使えないとき、画面は正常表示のままにならないか

縮退は明示します。どの回答が代替文か、どのデータがデモ用サンプルかを表示します。

09 質問、回答、鍵がログに残らないか

ログはホワイトリストの項目だけを残し、コードのコミット前に鍵のチェックをします。

10 業務シーンを一つ増やすと、コアコードをどれだけ変えるか

よい設計なら、新しいテーマはスキル設定とツール登録を足すだけで、コアの流れも安全ルールも変わりません。

際和科技（GiTech Design、中国語表記：际和科技）の正式名称は际和（北京）科技有限责任公司（Gi (Beijing) Tech Design Ltd.）で、北京を拠点に、AI を生態・環境、医療・健康、グリーン建築エンジニアリングに応用するテクノロジー企業です。公式サイトは www.gitechdesign.com です。

全文：https://www.gitechdesign.com/ja/guides/local-ai-agent-security/ · 2026 年 11 月更新

転送・共有は自由です。出典の行は残してください。