

人間による内容確認済み | 所見採用は保留中 | 実行検査なし

ツールごとに必須の制御経路を定める

Anthropic Claude Agent SDK 制御設計レビュー

公開情報を対象に弊会が独自に作成したレビューサンプル候補。対象企業・対象プロジェクトから依頼・承認・認定を受けたものではありません。

レビュー情報	対象
レビューID / 版	CRS-PUB-CASDK-HOOKS-20260929-001 / v0.1
方法 / レビュー日	MIXED_STATIC / 2026-09-30
Repository / commit	anthropics/claude-agent-sdk-python @ 37422c21d89fce79081d4f4fc59db6f7dce7ff72
Path scope	README.md / examples/hooks.py / src/claude_agent_sdk/types.py

EXECUTIVE SUMMARY

`allowed_tools` は自動承認の設定であり、Toolの利用範囲を限定する設定とは役割が異なります。`can_use_tool` は全呼出しに必ず入る検問ではありません。公式の型説明と例に従い、必要な検査をどのToolに必須とするか、変更後入力・子Agent・設定版・別経路まで含めて設計する必要があります。実際の実行順序や阻止結果は未観測です。

保護する作用とレビュー上の問い

ファイル変更・コマンド・MCPなどの作用を伴うTool実行。採用先の許可された作用と到達先を別途固定する。
利用可能なTool、自動承認、権限判断、`PreToolUse` を、作用ごとの必須制御としてどう組み合わせるか。

主な所見候補

A-F1	利用可能範囲と自動承認を分ける	採用保留
A-F2	必須検査を通る経路を固定する	採用保留
A-F3	識別子と最終入力を結び付ける	採用保留
A-F4	子Agentの帰属と権限範囲を定める	採用保留
A-F5	設定と判断の由来を固定する	採用保留
A-F6	同じ作用へ届く別Toolも確認する	採用保留

状態欄は所見採用の保留を示します。根拠・推論・主張範囲は人間確認済みで、実行試験の可否や脆弱性判定ではありません。

人間による内容確認済み | 所見採用は保留中 | 実行検査なし

実装へ渡す条件と検証の問い

以下は人間確認済みの制御条件案です。所見採用は保留中で、既存実装の欠陥認定や実行検査結果ではありません。

制御条件案	実装後に確かめる問い
A-F1 利用可能Tool、禁止Tool、自動承認、都度承認を別々の設定項目として固定する。	自動承認リストにないToolは、採用方針どおりに禁止または都度承認になるか。
A-F2 各作用に必須検査を割り当て、自動承認・allow 応答を含む経路でも省略されないことを確認する。	Bashが自動承認される設定でも、禁止入力の詳細検査が作用前に適用されるか。
A-F3 セッション・Tool・tool_use_id・Agent 帰属・最終入力を結び付ける。意味変更後の入力は必須条件を再評価する。	検査後に updatedInput 等で宛先やパスを変えても、変更後入力が必要な制御を通るか。
A-F4 主Agent と子Agent を明示的に区別し、セッション・Agent・呼出しごとに権限と判断を結び付ける。不明な帰属は権限を広げない。	子Agent B の要求に、子Agent A または親の承認が誤って使われないか。
A-F5 SDK/CLI 版・実効設定源・方針版・フック集合を記録し、競合・例外・時間切れ・ask/defer の扱いを明文化する。	設定源やフックの完了順が変わっても、同じ必須制御条件が維持されるか。
A-F6 作用を基準にToolと到達先を列挙する。利用可能な別経路は同じ制御条件で検査するか、利用を禁止する。	一つのToolで禁止した作用へ、許可された別Toolから到達できるか。

主張上限

公開情報を対象に、実案件で使用するレビュー構造・成果物形式に沿って作成した公開サンプル候補。根拠と主張の対応・推論・主張範囲は人間による内容確認済み。所見採用は保留中であり、静的レビューを実行結果・認証・実行許可・公開許可へ昇格させない。

人間による内容確認済み | 所見採用は保留中 | 実行検査なし

対象ソースと根拠

対象版はmoving ref ではなくexact commit で固定。行番号は取得したUTF-8 原本の始まりです。

- A-S1 L59-65

README.md / Using Tools

allowed_tools は自動承認リスト。未掲載Toolは permission_mode / can_use_tool 側の判断へ進むと説明。Tool集合から削除する設定ではない。
- A-S2 L46-70

examples/hooks.py / check_bash_command

Bash入力に指定文字列を検出すると PreToolUse の deny を返す。
- A-S3 L156-185

examples/hooks.py / example_pretooluse

Bashの自動承認設定と、Bashを対象にしたPreToolUse を同時に構成。実行を指示する例があるが本レビューでは未実行。
- A-S4 L223-245

src/claude_agent_sdk/types.py / ToolPermissionContext

tool_use_id と子Agent 文脈の agent_id を定義。
- A-S5 L314-339

src/claude_agent_sdk/types.py / PreToolUseHookInput

Tool名・入力・tool_use_id と任意の子Agent 帰属フィールドを定義。
- A-S6 L437-444

src/claude_agent_sdk/types.py / PreToolUseHookSpecificOutput

allow / deny / ask / defer と updatedInput を定義。
- A-S7 L2158-2186

src/claude_agent_sdk/types.py / can_use_tool / hooks

既許可Toolは can_use_tool を呼ばないと説明。PreToolUse のallow も同callback を省略。同一イベントの複数matcher は並行dispatch と説明。
- A-S8 L1974-1989

src/claude_agent_sdk/types.py / tools / allowed_tools

tools は利用可能な組込みTool集合、allowed_tools は自動承認を指定。
- A-S9 L2297-2306

src/claude_agent_sdk/types.py / setting_sources

未指定時は全設定源を読み、空配列でファイル設定を無効化するとの説明。
- A-S10 L2063-2067

src/claude_agent_sdk/types.py / disallowed_tools

禁止Toolはモデル文脈から除かれ、他で許可されても利用できないとの説明。
- A-S11 L258-264

src/claude_agent_sdk/types.py / PermissionResultAllow

許可応答の updated_input と updated_permissions を定義。

人間による内容確認済み | 所見採用は保留中 | 実行検査なし

所見と制御条件の対応

各行の根拠と推論を分け、所見採用は保留中のままです。

A-F1 利用可能範囲と自動承認を分ける

Fact: README は `allowed_tools` を自動承認リストと説明する。型には `tools` と `disallowed_tools` もある。

Inference: `allowed_tools` だけで利用可能Toolを限定したと扱うと、設計上の許容範囲を誤読し得る。

A-F2 必須検査を通る経路を固定する

Fact: 型説明では既許可ToolとPreToolUse のallow は `can_use_tool` を省略する。例はBashの自動承認とPreToolUse のdenyを組み合わせる。

Inference: `can_use_tool` 内だけに全件必須検査を置く設計は、記載された適用範囲と整合しない。

A-F3 識別子と最終入力を結び付ける

Fact: 入力型に `tool_use_id` があり、権限許可には `updated_input`、PreToolUse 出力には `updatedInput` がある。

Inference: 識別子フィールドの存在は内容不変の証明ではない。更新入力を許す場合、検査した入力と実行する入力の対応を採用設計で定める必要がある。

A-F4 子Agentの帰属と権限範囲を定める

Fact: ToolPermissionContext に子Agent の `agent_id` があり、フック入力にも任意の帰属フィールドがある。型説明は並行子Agent のイベントが交錯するとする。

Inference: 帰属フィールドだけから、親子間の権限継承・隔離が成立しているとは判断できない。

A-F5 設定と判断の由来を固定する

Fact: 型説明では設定由来の許可も `can_use_tool` を省略し得る。`setting_sources` 未指定時は全設定源を読む。同一イベントの複数matcher は並行dispatch と説明。

Inference: アプリ設定だけを記録しても実効方針を再現できない場合がある。複数フックの逐次依存も前提にできない。

A-F6 同じ作用へ届く別Toolも確認する

Fact: 例の `matcher` はBashを対象とし、`callback` はコマンド中の指定文字列を調べる。これは特定Tool・入力条件の例である。

Inference: 同じファイル変更や外部送信へ別Toolから到達できる採用構成では、その別経路も制御対象にする必要がある。

人間確認済みなのは根拠と主張の対応・推論・主張範囲です。正式所見採用、実行検査、安全性保証、認証、実行許可ではありません。

人間による内容確認済み | 所見採用は保留中 | 実行検査なし

A-F1 / A-F2

A-F1 利用可能範囲と自動承認を分ける

事実	READMEは <code>allowed_tools</code> を自動承認リストと説明する。型には <code>tools</code> と <code>disallowed_tools</code> もある。
推論	<code>allowed_tools</code> だけで利用可能Toolを限定したと扱うと、設計上の許容範囲を誤読し得る。
代替説明	採用先が <code>tools</code> ・ <code>disallowed_tools</code> ・ OS権限で別途限定している可能性がある。
必要条件案	利用可能Tool、禁止Tool、自動承認、都度承認を別々の設定項目として固定する。
将来の質問	自動承認リストにないToolは、採用方針どおりに禁止または都度承認になるか。

A-F2 必須検査を通る経路を固定する

事実	型説明では既許可ToolとPreToolUse のallow は <code>can_use_tool</code> を省略する。例はBashの自動承認とPreToolUse のdenyを組み合わせる。
推論	<code>can_use_tool</code> 内だけに全件必須検査を置く設計は、記載された適用範囲と整合しない。
代替説明	全件検査をPreToolUse や実行アダプターに置く設計なら、この不整合を回避し得る。
必要条件案	各作用に必須検査を割り当て、自動承認 ・ allow 応答を含む経路でも省略されないことを確認する。
将来の質問	Bashが自動承認される設定でも、禁止入力 of 必須検査が作用前に適用されるか。

人間による内容確認済み | 所見採用は保留中 | 実行検査なし

A-F3 識別子と最終入力を結び付ける

内容確認済み / 所見採用は保留中 / 根拠: A-S4, A-S5, A-S6, A-S11

ソース事実

入力型に `tool_use_id` があり、権限許可には `updated_input`、PreToolUse 出力には `updatedInput` がある。

推論と所見候補

識別子フィールドの存在は内容不変の証明ではない。更新入力を許す場合、検査した入力と実行する入力の対応を採用設計で定める必要がある。

代替説明

CLIやアダプターに再検査がある可能性がある。対象型定義だけでは実行入力への反映手順を断定できない。

必要な制御条件案

セッション・Tool・`tool_use_id`・Agent 帰属・最終入力を結び付ける。意味変更後の入力は必須条件を再評価する。

将来の検証質問

検査後に `updatedInput` 等で宛先やパスを変えても、変更後入力が必要な制御を通るか。

主張できないこと

`tool_use_id` が全環境で一貫または入力改変を防ぐという主張。

未観測範囲

変更・並行処理・再開時のCLI実装。

人間による内容確認済み | 所見採用は保留中 | 実行検査なし

A-F4 子Agentの帰属と権限範囲を定める

内容確認済み / 所見採用は保留中 / 根拠: A-S4, A-S5, A-S7

ソース事実

ToolPermissionContext に子Agent の agent_id があり、フック入力にも任意の帰属フィールドがある。型説明は並行子Agent のイベントが交錯するとする。

推論と所見候補

帰属フィールドだけから、親子間の権限継承・隔離が成立しているとは判断できない。

代替説明

採用先がAgent 別のTool集合や実行環境を分離している可能性がある。

必要な制御条件案

主Agent と子Agent を明示的に区別し、セッション・Agent ・呼出しごとに権限と判断を結び付ける。不明な帰属は権限を広げない。

将来の検証質問

子Agent B の要求に、子Agent A または親の承認が誤って使われないか。

主張できないこと

agent_id が本人性を認証するという主張。

未観測範囲

権限継承実装、子Agent 設定解決、同時実行。

人間による内容確認済み | 所見採用は保留中 | 実行検査なし

A-F5 設定と判断の由来を固定する

内容確認済み / 所見採用は保留中 / 根拠: A-S7, A-S9, A-S6

ソース事実

型説明では設定由来の許可も `can_use_tool` を省略し得る。`setting_sources` 未指定時は全設定源を読む。同一イベントの複数`matcher` は並行`d` `ispatch` と説明。

推論と所見候補

アプリ設定だけを記録しても実効方針を再現できない場合がある。複数フックの逐次依存も前提にできない。

代替説明

実行環境が設定を封じ、フックを独立に設計しているなら影響は限定される。

必要な制御条件案

SDK/CLI 版・実効設定源・方針版・フック集合を記録し、競合・例外・時間切れ・`ask/defer` の扱いを明文化する。

将来の検証質問

設定源やフックの完了順が変わっても、同じ必須制御条件が維持されるか。

主張できないこと

並行フック自体が欠陥であるという主張。

未観測範囲

複数判断の統合規則、全失敗経路、配布CLI。

人間による内容確認済み | 所見採用は保留中 | 実行検査なし

A-F6 同じ作用へ届く別Toolも確認する

内容確認済み / 所見採用は保留中 / 根拠: A-S2, A-S3, A-S8, A-S10

ソース事実

例の `matcher` はBashを対象とし、`callback` はコマンド中の指定文字列を調べる。これは特定Tool・入力条件の例である。

推論と所見候補

同じファイル変更や外部送信へ別Toolから到達できる採用構成では、その別経路も制御対象にする必要がある。

代替説明

別Toolが利用不可、またはOS・MCPサーバー側で同じ作用を阻止する構成もあり得る。

必要な制御条件案

作用を基準にToolと到達先を列挙する。利用可能な別経路は同じ制御条件で検査するか、利用を禁止する。

将来の検証質問

一つのToolで禁止した作用へ、許可された別Toolから到達できるか。

主張できないこと

文字列例の迂回を実証した、または全経路を網羅したという主張。

未観測範囲

全Tool・MCP実装・外部接続・OS権限。

人間による内容確認済み | 所見採用は保留中 | 実行検査なし

主張上限と次工程

未観測範囲と成果物の来歴

- ・ソースを静的に読んだ結果であり、対象SDKやサンプルを実行していない。
- ・実顧客の採用済み方針、稼働環境、非公開資料、運用ログは受領していない。
- ・必要条件は弊会の提案。対象プロジェクトが同じ業務要件を約束しているとの主張ではない。
- ・最新バージョンや製品全体への一般化、脆弱性・安全性・適合性の判定は行わない。

人間による内容確認

根拠と主張の対応、推論・代替説明、主張上限を人間確認済み。所見採用は保留中で、正式所見・実行検査結果・安全性保証・認証・実行許可には昇格していません。確認時刻はCRS記録に保持します。

AI支援と助言関与

OpenAI / ChatGPT Work 上のAIが資料読解、所見候補、改善条件案、文書生成と体裁確認を支援。作成と助言は同一工程で、独立第三者レビューではありません。

主張上限

公開情報を対象に、実案件で使用するレビュー構造・成果物形式に沿って作成した公開サンプル候補。根拠と主張の対応・推論・主張範囲は人間による内容確認済み。所見採用は保留中であり、静的レビューを実行結果・認証・実行許可・公開許可へ昇格させない。