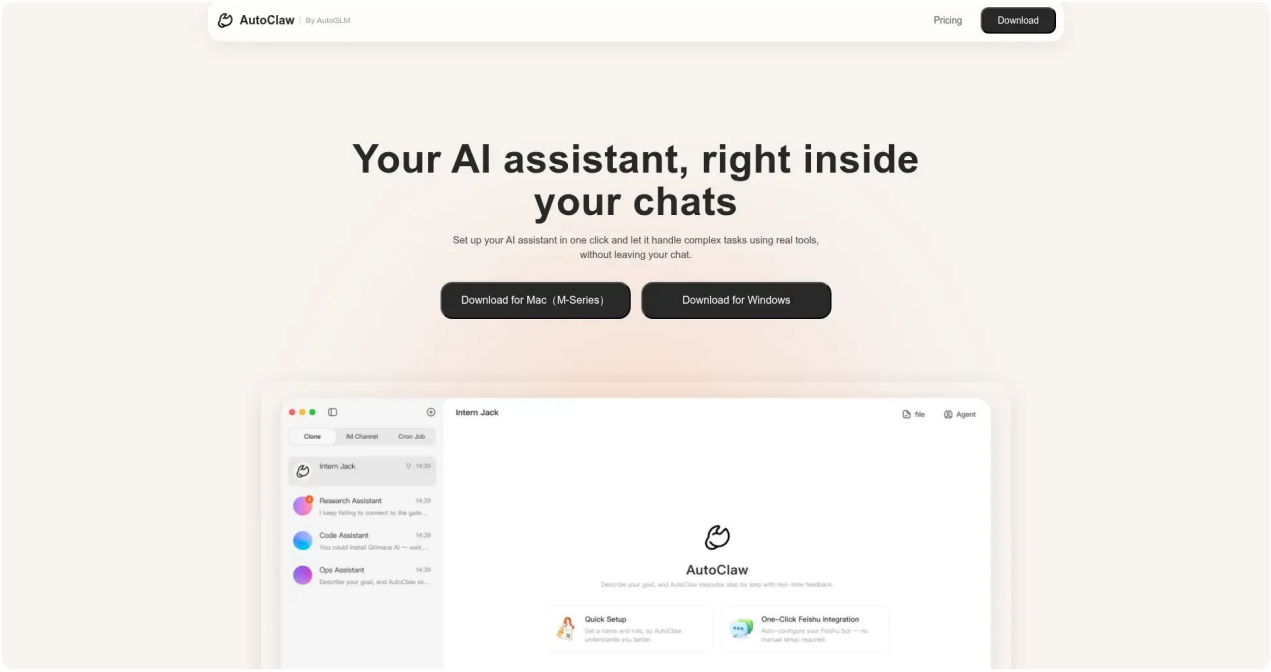

AutoGLM

Agente de inteligencia artificial multimodal diseñado para la navegación autónoma de interfaces digitales en móviles y navegadores. Permite a desarrolladores, ingenieros de RPA y departamentos de innovación automatizar tareas complejas mediante la interpretación visual de la GUI, emulando el comportamiento humano sin depender de APIs oficiales. Es ideal para profesionales que gestionan flujos de trabajo multietapa y necesitan extraer datos o ejecutar acciones en aplicaciones de terceros.

Visitar Sitio Oficial Preguntar a ChatGPT Preguntar a Claude Preguntar a Grok

Contenido del Dossier

INFORMACIÓN DE LA HERRAMIENTA

Qué y para quién es
AutoGLM es un agente de inteligencia artificial multimodal diseñado para la navegación autónoma de interfaces digitales, capaz de ejecutar tareas complejas en dispositivos móviles y navegadores web emulando el comportamiento humano. Está dirigido principalmente a desarrolladores de software, ingenieros de automatización de procesos (RPA), departamentos de innovación tecnológica y profesionales que buscan delegar tareas repetitivas de gestión de aplicaciones. En el ámbito profesional, es una herramienta disruptiva para sectores que dependen intensamente de la gestión de datos en aplicaciones de terceros donde no existen APIs oficiales.

Principal ventaja profesional
En mi opinión profesional, tras analizar su capacidad de comprensión visual, la ventaja definitiva es su capacidad de razonamiento "zero-shot" sobre interfaces desconocidas. A diferencia de las herramientas de automatización tradicionales basada en selectores o coordenadas fijas, AutoGLM entiende visualmente qué es un botón de "comprar" o un campo de "búsqueda" independientemente del diseño, lo que reduce drásticamente los costes de mantenimiento de scripts de automatización.

Para quién no es
Como profesional, considero que no es apto para empresas con políticas de privacidad extremadamente rígidas o sectores donde el procesamiento de capturas de pantalla en la nube suponga un riesgo de cumplimiento legal (como banca suiza o defensa), ya que el modelo requiere "ver" la interfaz para operar. Tampoco es para usuarios finales que busquen un asistente de voz básico, ya que su potencial reside en la ejecución de flujos de trabajo multietapa.

Funcionalidades clave
- Navegación autónoma en smartphones y navegadores web mediante interpretación visual de la GUI.
- Planificación de tareas de varios pasos: Puede recibir una orden compleja (ej. "pide un café similar al de ayer") y desglosarla en acciones dentro de aplicaciones comerciales.
- Capacidad de corrección de errores en tiempo real: Si la aplicación muestra un mensaje inesperado, el agente intenta rutas alternativas.
- Comprensión multimodal avanzada basada en el modelo ChatGLM, permitiendo procesar texto e imagen simultáneamente.
- Simulación de gestos humanos como clics, desplazamientos y escritura de texto en campos de entrada.

Precios
- Versión gratuita: Actualmente disponible en fases de prueba abierta y demostración técnica para desarrolladores, sujeta a límites de uso de inferencia por parte de Zhipu AI.
- Rango de precios: Basado en consumo de tokens o por llamada a la API (modelo pago por uso).
- Versiones de pago: Planes empresariales que priorizan la latencia de respuesta y ofrecen mayor capacidad de concurrencia para procesos automatizados masivos.

Perfil del usuario
- Empresas de comercio electrónico que requieren monitorización de precios y competencia en apps móviles.
- Departamentos de QA y Testing de software para pruebas de regresión automatizadas.
- Consultoras de transformación digital que implementan soluciones de hiperautomatización.
- Profesionales de marketing para la gestión automatizada de interacciones en redes sociales.

Nivel técnico requerido
- Nivel técnico requerido para su uso: Medio. Requiere saber estructurar prompts detallados y entender el flujo lógico de las aplicaciones.
- Nivel técnico requerido para su instalación/configuración: Alto. Es necesario conocimiento en integración de APIs, gestión de entornos Python y configuración de herramientas de control remoto de dispositivos (ADB o similares).
- Necesidades de soporte: Requiere supervisión de ingenieros de IA para el ajuste de los parámetros de temperatura y tokens de salida.
- Conocimientos necesarios: Programación en Python, manejo de JSON y familiaridad con modelos de lenguaje multimodales (LMM).

Ejemplos de uso profesional
- Automatización de informes: Extraer datos de aplicaciones móviles corporativas que no tienen versión web ni exportación de datos.
- Atención al cliente: Delegar al agente la búsqueda de estados de pedido en apps logísticas de terceros para responder a tickets de soporte.
- Gestión de RRHH: Publicación automatizada de ofertas en múltiples portales de empleo de forma simultánea navegando por cada interfaz.

Uso y distribución
- Versión web: Disponible mediante SDK para integración en navegadores.
- Versión móvil: Integración nativa en sistemas operativos específicos mediante capas de accesibilidad.
- CLI: Herramientas de línea de comandos para desarrollo y pruebas.

Integraciones
- Facilidad de integración: Full code. Requiere desarrollo para conectar los disparadores de negocio con el agente.
- API propia: Dispone de API REST para el envío de capturas de pantalla y recepción de comandos de acción.
- Descripción de integraciones: Se integra nativamente con el ecosistema de Zhipu AI y puede conectarse mediante middleware a herramientas como Zapier o Make para disparar acciones basadas en eventos externos.

Notas finales
Veredicto técnico
He verificado que AutoGLM representa un salto cualitativo sobre el RPA tradicional; lo considero una herramienta de gran utilidad para empresas que han tocado techo con la automatización basada en reglas. No es una solución "instalar y usar", requiere una infraestructura técnica sólida detrás, pero la eficiencia que aporta al eliminar la necesidad de programar cada movimiento del ratón compensa con creces el esfuerzo de implementación inicial.

Información legal, licencias, contratos
El uso de AutoGLM está sujeto a los términos de servicio de Zhipu AI. Es fundamental revisar la política de tratamiento de imágenes, ya que el sistema captura visualmente la pantalla del dispositivo durante la ejecución de las tareas para poder "entender" la interfaz.

Otros
Es importante destacar que el rendimiento de la herramienta está condicionado a la latencia de red, ya que el procesamiento visual es intensivo en datos.

Fuentes consultadas:
- Sitio web oficial: https://autoglm.z.ai
- Repositorio oficial de modelos: https://github.com/THUDM/ChatGLM
- Documentación técnica Zhipu AI: https://open.bigmodel.cn
- Comunidad de investigación: https://keg.cs.tsinghua.edu.cn

CONSEJOS DE IMPLANTACIÓN

Aplicación profesional
En mi opinión profesional, AutoGLM es la pieza que faltaba para cerrar la brecha de la hiperautomatización en empresas que dependen de ecosistemas cerrados o aplicaciones móviles sin API. Según mi experiencia, es ideal para compañías con procesos de back-office intensivos en el uso de herramientas SaaS externas o aplicaciones legacy. El presupuesto necesario no es solo el coste de inferencia (pago por uso de tokens), sino la inversión en infraestructura de servidores o dispositivos móviles que actuarán como "hosts" del agente. Lo que más me gusta es su capacidad para romper la fragilidad del RPA tradicional: si una aplicación web actualiza su interfaz y cambia un botón de sitio, AutoGLM no se rompe, simplemente lo vuelve a encontrar visualmente.

Madurez digital requerida
- **Usuarios y equipo:** El equipo operativo debe estar familiarizado con la supervisión de sistemas autónomos ("human-in-the-loop"). No basta con saber usar la aplicación, deben entender cómo definir objetivos claros para una IA.
- **Empresa y departamentos:** Se requiere una madurez alta. Es necesario un departamento de IT o DevOps capaz de gestionar entornos de ejecución (como contenedores Docker o granjas de móviles mediante ADB) y una gobernanza de datos clara, ya que el agente "verá" información sensible en pantalla.

Plan orientativo de implantación
Pasos necesarios y estimaciones
- **Tiempos estimados de despliegue:** De 4 a 8 semanas para un flujo de trabajo estable en entornos de producción.
- **Evaluación inicial (Semana 1-2):** Auditoría de los procesos manuales candidatos. Al usarlo te das cuenta de que no todos los procesos son óptimos para un agente visual; hay que priorizar aquellos donde no exista API y el volumen de tareas sea alto.
- **Configuración técnica y POC (Semana 3-4):** Configuración de los modelos de Zhipu AI, preparación de los entornos de ejecución (emuladores o dispositivos físicos) y primera prueba de concepto en un entorno controlado (sandbox).
- **Refinamiento de prompts y lógica (Semana 5-6):** Ajuste de las instrucciones del sistema para minimizar alucinaciones visuales y asegurar que los "clics" se realizan en los elementos correctos.
- **Piloto y supervisión (Semana 7-8):** Despliegue con supervisión humana constante para validar la tasa de éxito y ajustar parámetros de latencia.

Necesidades de formación del equipo
- Capacitación en ingeniería de prompts multimodales (cómo describir tareas visuales).
- Formación en gestión de excepciones: qué hacer cuando el agente se queda en un bucle o no interpreta correctamente un pop-up.
- Conocimientos básicos de arquitectura de agentes autónomos para los responsables de procesos.

Perfiles necesarios
- **Perfiles técnicos necesarios:** Ingenieros de IA/ML para el ajuste del modelo, desarrolladores Python y especialistas en QA Automation.
- **Personal externo recomendado:** Consultores expertos en ética de IA y privacidad de datos para asegurar el cumplimiento legal de las capturas de pantalla.
- **Otros:** Un "Process Owner" que conozca perfectamente el flujo de negocio que se va a automatizar.

Retorno de la inversión (ROI)
- **Tiempos:** El ahorro de tiempo es casi inmediato tras la estabilización (reducción del 70-90% del tiempo manual). El retorno de la inversión económica suele verse entre los 6 y 12 meses.
- **Cómo medirlo (KPIs):** Tasa de éxito de la tarea (Task Success Rate), tiempo medio de ejecución comparado con un humano, y especialmente el coste de mantenimiento de los scripts en comparación con el RPA basado en selectores HTML.

Otros
Mi experiencia en implantaciones me lleva a pensar que el mayor cuello de botella de AutoGLM será la latencia de red y el tiempo de inferencia del modelo visual. Es vital realizar pruebas de carga para entender cuántos agentes concurrentes puede soportar tu infraestructura sin que la respuesta del modelo degrade

la experiencia del proceso de negocio. Además, recomiendo establecer "zonas ciegas" o máscaras de privacidad para ocultar datos sensibles (como contraseñas o datos bancarios) antes de que la captura de pantalla sea enviada al modelo para su procesamiento.

TUTORIAL BÁSICO

Instalación
AutoGLM es un framework de agentes multimodales diseñado para controlar dispositivos móviles mediante lenguaje natural.
- **Requisitos de Sistema**: Se recomienda Python 3.10 o superior. Es compatible con Android 7.0+, HarmonyOS Next e iOS (mediante WebDriverAgent).
- **Herramientas de Depuración**: Es indispensable instalar ADB para Android o HDC para dispositivos Huawei (HarmonyOS), asegurándose de que el binario esté en el PATH del sistema.
- **Configuración del Dispositivo**: Debes activar las "Opciones de Desarrollador" y habilitar "Depuración por USB". En muchos terminales Android, también es necesario activar "Instalar vía USB" y "Depuración USB (ajustes de seguridad)" para permitir la simulación de toques.
- **ADB Keyboard**: Es fundamental instalar el APK de ADB Keyboard en el teléfono y seleccionarlo como método de entrada predeterminado para que el agente pueda escribir texto correctamente.
- **Checklist de Instalación**:
- [ ] pip install -r requirements.txt && pip install -e . ejecutado.
- [ ] Dispositivo reconocido mediante adb devices.
- [ ] API Key de BigModel (Zhipu AI) o servidor local (vLLM/SGLang) configurado.

Uso en el día a día
Según mi experiencia, AutoGLM destaca por su capacidad de planificar tareas complejas cruzando diferentes aplicaciones.
- **Modo Interactivo**: Utiliza python main.py --base-url <API_URL> --model <MODEL_NAME> para iniciar una sesión donde puedes encadenar comandos.
- **Automatización de aplicaciones**: Puedes pedir tareas como "Busca una cafetería en Google Maps y envíale la ubicación a mi contacto en WhatsApp". El agente se encarga de cambiar de app automáticamente.
- **Confirmación Sensible**: Lo que más me gusta es el sistema de seguridad; para acciones críticas (como pagos o borrado de datos), el agente solicita confirmación humana mediante un callback.
- **Intervención Humana (Takeover)**: Si encuentras un CAPTCHA o un login biométrico, el sistema pausará y te pedirá que tomes el control manualmente antes de continuar.

Trucos de experto
- **Depuración Inalámbrica**: Una vez conectado por USB, usa adb tcpip 5555 y luego adb connect <IP_IP_DEL_MOVIL>:5555. Esto permite usar el agente sin cables, lo cual es mucho más cómodo para pruebas largas.
- **Optimización de Prompts**: Si el agente falla en una app específica, puedes modificar los archivos en phone_agent/config/prompts_zh.py (o en.py) para añadir instrucciones específicas sobre cómo navegar en esa interfaz.
- **Modo Verbose**: Al usarlo te das cuenta de que activar el modo --verbose es vital. Te permite ver el "razonamiento" o la cadena de pensamiento (<think>) del modelo antes de ejecutar la acción, lo que ayuda a entender por qué ha fallado un paso.
- **Integración con Midscene.js**: Para desarrolladores web, AutoGLM se puede integrar con Midscene para automatizar flujos que saltan del navegador a la interfaz nativa del móvil.

Posibles problemas/incidencias
- **Latencia de Red**: Dado que envía capturas de pantalla al modelo visual para su análisis, una conexión lenta entre el PC y la API (o el móvil) degradará significativamente la experiencia.
- **Incompatibilidad de Entrada**: Si el agente intenta escribir y no pasa nada, el 90% de las veces es porque no se ha activado el ADB Keyboard o no se han concedido permisos de entrada en los ajustes de desarrollador.
- **Errores de Resolución**: En mi opinión profesional, el modelo funciona mejor en resoluciones estándar. Si usas pantallas plegables o tablets con ratios de aspecto extraños, el agente puede fallar al calcular las coordenadas de los clics (Tap).
- **Check de Despliegue**: Si usas un servidor local, utiliza siempre el script scripts/check_deployment_cn.py para verificar que el modelo visual está respondiendo con el formato JSON correcto antes de intentar controlar el móvil.

Otros
- **Modelos Disponibles**: Existen versiones de 9B parámetros tanto en chino como multilingüe. El modelo multilingüe es preferible si utilizas aplicaciones con interfaz en inglés o español.

- **Plataformas de Inferencia**: Además de BigModel oficial, puedes usar infraestructuras como ModelScope o Novita AI para ejecutar el backend de AutoGLM.

PREGUNTAS FRECUENTES

¿Qué es AutoGLM y en qué se diferencia de los sistemas RPA tradicionales?

AutoGLM es un agente de inteligencia artificial multimodal que utiliza modelos de lenguaje visual (VLM) para navegar de forma autónoma por interfaces digitales. A diferencia del RPA tradicional, que depende de reglas rígidas y selectores de código, AutoGLM emplea razonamiento 'zero-shot' para interpretar visualmente elementos de la interfaz, como botones o formularios, permitiéndole operar en aplicaciones desconocidas o con diseños cambiantes sin necesidad de reprogramación.

¿Para qué tipo de tareas profesionales está diseñado?

Está diseñado para la ejecución de flujos de trabajo multietapa que requieren interacción con interfaces gráficas de usuario (GUI) en dispositivos móviles y navegadores web. Sus aplicaciones incluyen la extracción de datos en plataformas sin API oficial, la automatización de pruebas de regresión (QA), la monitorización de precios en tiempo real y la gestión operativa en herramientas de terceros donde no es posible la integración directa.

¿Cuál es el coste del servicio y su disponibilidad?

El modelo opera bajo un esquema de pago por uso basado en el consumo de tokens o por llamadas a la API de Zhipu AI. Actualmente existe una versión de demostración técnica y fases de prueba abierta con límites de inferencia gratuitos. Para implementaciones corporativas, se ofrecen planes empresariales con prioridad en la latencia y mayor capacidad de concurrencia.

¿Es posible descargar el código desde GitHub?

Si bien el ecosistema ChatGLM en el que se basa tiene presencia en repositorios como los de THUDM en GitHub, AutoGLM se distribuye principalmente como una solución de servicio a través de SDKs y APIs de Zhipu AI para integración en entornos de desarrollo profesionales y control de dispositivos.

¿Qué nivel de conocimientos técnicos se requiere para su implementación?

La configuración e instalación exigen un perfil técnico alto, con sólidos conocimientos en programación Python, gestión de entornos de desarrollo, manejo de archivos JSON e integración de APIs REST. También es necesario dominar herramientas de control remoto de dispositivos, como Android Debug Bridge (ADB), para la ejecución de tareas en entornos móviles.

¿Cómo aborda AutoGLM la seguridad y la privacidad de los datos?

AutoGLM requiere capturar visualmente la pantalla del dispositivo para procesar e interpretar la interfaz en la nube. Esto implica que la información sensible visible durante la ejecución es procesada por los servidores de Zhipu AI. Por ello, no se recomienda para sectores con normativas de privacidad extremadamente estrictas o infraestructuras críticas donde el envío de capturas de pantalla externas suponga un riesgo de cumplimiento legal.

¿Cumple con la normativa española de protección de datos?

El cumplimiento de la normativa española y europea (RGPD) depende de los acuerdos de procesamiento de datos establecidos con Zhipu AI. Al ser una tecnología que procesa información visual en servidores remotos, las empresas deben realizar una evaluación de impacto de protección de datos (EIPD) antes de integrarlo en procesos que manejen datos de carácter personal.

¿Qué limitaciones técnicas presenta actualmente?

El rendimiento está condicionado por la latencia de red debido al alto volumen de datos que supone el procesamiento visual. Además, aunque tiene capacidad de corrección de errores, requiere supervisión técnica para ajustar parámetros como la 'temperatura' del modelo y asegurar que el razonamiento lógico sea consistente con los objetivos de negocio.

CONTRATOS Y CONDICIONES

Opinión inicial
Tras verificar los contratos y las condiciones de Zhipu AI para AutoGLM, mi opinión técnica es que nos encontramos ante una herramienta de **alto impacto legal y riesgo de cumplimiento elevado** para una empresa española. Aunque tecnológicamente es disruptiva al actuar como un "operador humano" sobre interfaces, su arquitectura se basa en el envío constante de capturas de pantalla a servidores externos para su análisis. Según los documentos consultados, la responsabilidad del cumplimiento recae casi íntegramente en el desarrollador/empresa que implementa la solución, descargando al proveedor de cualquier mal uso. La legislación aplicable predominante es la de China (sede de Zhipu AI), lo que genera un conflicto directo con las garantías de soberanía de datos de la UE.

Principales recomendaciones
- **Priorizar despliegue local**: Si la empresa cuenta con infraestructura, optar por el despliegue "Local Model + Local Execution" para evitar que las capturas de pantalla (que pueden contener datos sensibles de clientes o empleados) salgan de la red corporativa.
- **Implementar confirmación humana**: Establecer obligatoriamente un paso de validación manual para operaciones críticas (pagos, gestión de derechos o acceso a salud) tal como sugieren los términos de uso del fabricante.
- **Auditoría de pantallas**: Configurar filtros de privacidad que pixelan o bloqueen áreas con datos personales (como DNI o tarjetas) antes de que el agente procese la imagen.
- **Limitación de permisos**: Usar el principio de "mínimo privilegio" en las credenciales de ADB o accesibilidad que el agente utiliza en los dispositivos.

Ley de Inteligencia Artificial (AI Act)
- **Clasificación**: Podría considerarse de **Alto Riesgo** si se utiliza para la gestión de recursos humanos, acceso a servicios públicos esenciales o evaluación crediticia, debido a su capacidad de tomar decisiones autónomas en interfaces críticas.
- **Transparencia**: El sistema debe informar claramente a cualquier usuario que esté interactuando con un agente automatizado.
- **Supervisión humana**: La herramienta requiere por diseño mecanismos de intervención inmediata para evitar comportamientos no deseados en "bucle".

Privacidad y protección de datos
- **Responsabilidades**: Zhipu AI actúa solo como proveedor del marco de trabajo. La empresa española es la **Responsable del Tratamiento** y debe realizar una Evaluación de Impacto (EIPD) antes de su despliegue.
- **Ubicación de los datos**: Si se usa la versión Cloud, los datos se procesan en infraestructuras fuera del Espacio Económico Europeo (principalmente China).
- **Transferencia internacional**: Existe una transferencia internacional de datos de alto riesgo. Sin una Decisión de Adecuación con China, esto requiere Cláusulas Contractuales Tipo (SCC) y medidas suplementarias de cifrado extremo a extremo.
- **Derechos ARCO**: Es técnicamente complejo garantizar el derecho de supresión o acceso sobre datos capturados en logs de imágenes de entrenamiento o inferencia si no se gestionan localmente.

Propiedad intelectual
- **Propiedad de datos**: Los términos generales suelen indicar que el usuario retiene la propiedad de los datos de entrada, pero otorga licencias de uso para la mejora del modelo en versiones no empresariales.
- **Propiedad del resultado**: El código generado o las automatizaciones creadas bajo la licencia **Apache 2.0** (en su versión Open) permiten el uso comercial, pero los resultados de la ejecución dependen de la titularidad de las aplicaciones sobre las que opera.

Usos y prohibiciones
- **Usos prohibidos**: Manipulación de datos a gran escala (brushing), registros masivos fraudulentos, scraping que viole archivos robots.txt o términos de terceros, y cualquier actividad de "…abuso (…o de subsidios/cupones).
- **Usos admitidos**: Automatización de procesos internos (RPA), pruebas de software (QA), y navegación asistida para accesibilidad bajo supervisión.

Seguridad y certificaciones
- **Seguridad**: El uso de ADB (Android Debug Bridge) abre una superficie de ataque considerable; si el agente

es comprometido, el atacante tiene control total del dispositivo.
- **Certificaciones**: No se han verificado certificaciones específicas de Esquema Nacional de Seguridad (ENS) o ISO 27001 validadas por organismos europeos en su documentación oficial.

Otros
- **Conflictos de términos de terceros**: Al usar AutoGLM para navegar en apps de terceros (Instagram, LinkedIn, bancos), la empresa podría estar incumpliendo los términos de servicio de esas plataformas que prohíben explícitamente el uso de "bots" o agentes automatizados, lo que puede derivar en el bloqueo de cuentas corporativas.

Fuentes consultadas:
- Privacidad oficial Open-AutoGLM
- Licencia Apache 2.0 AutoGLM
- Condiciones de servicio Zhipu AI
- Documentación técnica ChatGLM