INTELIGENCIA ARTIFICIAL

El nuevo reto en la protección de datos

Glenda Suárez Cabrera
CISSP, CISA, CISM
Director Quality, Risk, Compliance & Security at Pitcher A.G. www.pitcher.com
Miembro ISACA Emerging Trends Working group

NOS PREGUNTAMOS…

LA REVOLUCION CHATGPT

"Tan revolucionaria como la creación de internet"

"Procesador de lenguaje natural más cercano a AGI"

"Ahora todo el mundo es programador, ahora todo el mundo es abogado, ahora todo el mundo es médico… ahora todo el mundo es lo que quiere con gpt"

Pánico Desatado

Home > News > Computing

Europol warns ChatGPT is being used to commit crime

By Sead Fadilpašić published 13 days ago

Countless ways to

Jailbreaking ChatGPT: how AI chatbot safeguards can be bypassed

TECH

Italy became the first Western country to ban ChatGPT. Here's what other countries are doing

All Open Letters

Pause Giant AI Experiments: An Open Letter

We call on all AI labs to immediately pause for at least 6 mo than GPT-4.

Signatures 20466

Add your signature

https://futureoflife.org/open-lett

Fake news, manipulación y hackeos: cómo ChatGPT es un "profundo riesgo para la Humanidad"

Musk, Wozniak y Harari encabezan un manifiesto firmado por miles de expertos que piden parar su desarrollo para regular y controlarla.

Samsung workers made a major error by using ChatGPT

By Lewis Maddison published 7 days ago

Samsung meeting notes and new source code are now in the wild after being leaked in ChatGPT

The danger of 'hallucinations'

*"Como se establece en los **Principios de IA de Asilomar**, la IA podría representar un **cambio profundo en la historia**, y debe planificarse y administrarse con cuidado. Desafortunadamente, este nivel de planificación no está ocurriendo…los laboratorios de IA han entrado en **una carrera fuera de control** para desarrollar mentes digitales cada vez más poderosas que **nadie, ni siquiera sus creadores, pueden entender. predecir o controlar de forma fiable**"*

HABLEMOS DE ETICA Y TRANSPARENCIA

"Los desafíos con Chat GPT son desafíos similares a los que vemos con los modelos básicos de lenguaje grande: puede inventar e los hechos"

"No es demasiado pronto para que los gobiernos intercedan. Es muy importante que todos comiencen a involucrarse, dado el impacto que van a tener estas tecnologías"

"La IA puede ser mal utilizada, o puede ser utilizada por malos actores. Entonces, hay preguntas sobre cómo gobernar el uso de esta tecnología a nivel mundial.

Mira Murati (CTO Open AI)

Fuente: Revista Time

"Alcanzar el AGI es una bonificación, no un requisito para el beneficio global"

Reid Hoffmann (Inversor en Open AI)
Fuente: Revista Forbes

…"el impacto de los modelos de IA debe debatirse ahora, dado que una vez ya lanzados no se podrán retirar. Es como una especia invasora…necesitamos políticas a la velocidad de la tecnología"

Aviv Ovadya (Investigador Hardvard)
Fuente: Revista Forbes

LOS NUEVOS AGENTES DE IA

La Carrera de Herramientas y Agentes basados en GPT se ha desatado…

Amazon respalda Stability AI.

Zapier integra Claude de Anthropic.

Bloomberg lanzó su propio LLM específicamente para finanzas

Midjourney lanzó un nuevo comando - ingeniería inversa a cualquier imagen como quieras. Toma la foto del Papa con la chaqueta blanca.

HuggingGPT: muestra la conexión de chatgpt con otros modelos en hugging face

Mckay Wrigley puede crear sitios web desde cero solo con sus órdenes de voz

Large Model Applications By Modality

Natural Language

- **Social Agents**: Character.AI, Woebot, Replika
- **Developer**: Cohere, Humanloop
- **Knowledge Mgt**
- **Writing**: HyperWrite, NovelAI, summari, grammarly, AI21labs
- **Search**: neeva, stealth, DASHWORKS, Metaphor
- **Legal**: stealth
- **Marketing Copy**: copy.ai, copysmith, Hypotenuse AI, anyword, Writesonic, Jasper
- **Support/Sales**: CRESTA, ada, GONG, Tymely
- **Voice/STT/TTS**: Otter.ai, RESEMBLE.AI, MURF, PolyAI, reSpeecher, WELLSAID

Games/Characters

inworld, AI DUNGEON, infinite story

Software Actions

Adept, bardeen, rby AI, stealth

Image

Midjourney, ROSEBUD.AI, reface, stability.ai, WOMBO, Enstil, craiyo

Music

SOUNDRAW, Aiva Technologies

Video

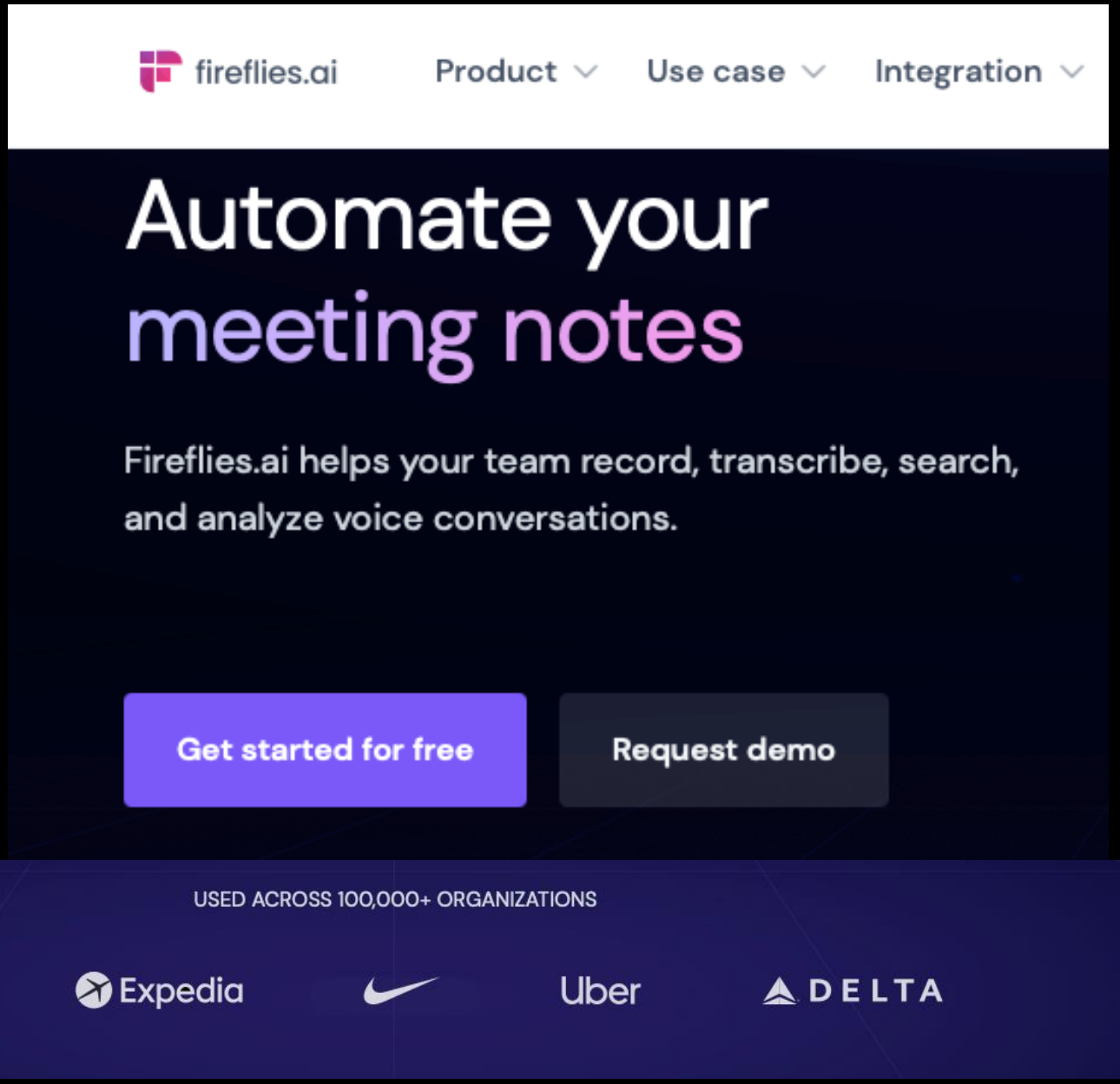
- **Generation**: RunwayML
- **Personal**: Synthesia, Hour One., tavus, EmbodyMe
- **Editing**: Reduct Video, deepdub.ai

Codegen

- **SQL / Data Model**: cogram, stealth
- **General**: tabnine, replit, stealth, MUTABLEAI, Magic
- **Documentation**: Mintlify, Stenography
- **App Design**: Debuild

CASO DE ESTUDIO: FIREFLIES.AI

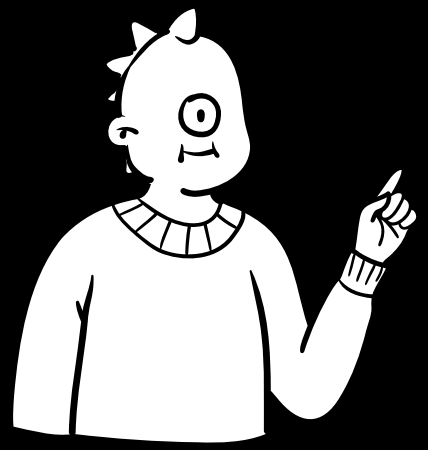
- Transcribe, resume y **analiza** reuniones y **conversaciones de voz**.
- Asigna tareas
- **Determina quién habló más** tiempo.
- Identifica si la **reunion fue positiva o negativa.** Si alguien estaba **enojado**.
- Identifica **competidores** y otros temas importantes.



Caso de uso:
Oportunidad para entender mejor las necesidades de nuestros clientes.
Potenciar ventas, evitar pérdida de cliente.
Eliminar la parte administrativa.

CASO DE ESTUDIO: FIREFLIES.AI

Alguien se ha mirado los T&Cs y La Política de Privacidad?



By making available any User Content through the Services, you hereby grant to Fireflies a worldwide, irrevocable, perpetual, non-exclusive, transferable, royalty-free license, with the right to sublicense, to use, access, view, copy, adapt, modify, distribute, license, sell, transfer, publicly display, publicly perform, transmit, stream, broadcast and otherwise exploit such User Content on, through or by means of the Services. We do not claim any ownership rights in any such User Content and nothing in this Agreement will be deemed to restrict any rights that you may have to use and exploit any such User Content.

The Services may provide voice, video and text chat, forum or bulletin board tools to users and the ability to create an "avatar" or other customized profile and Account information. Information that you provide through the use of these tools will be available to the public generally. Fireflies has no obligation to keep private any information that you disclose to other users or the public using these functions. You

USED ACROSS 100,000+ ORGANIZATIONS

NETFLIX Expedia Uber DELTA

IMPLICACION EN LA PROTECCION DE DATOS PERSONALES Y EL RGPD

La adopción de tecnologías de IA puede relentizar e incluso dar marcha atrás en el cumplimiento con el RGPD porque:

- ***Consentimiento o base legal:*** *No se puede garantizar transparencia total cuando el consentimiento es parcial, o falta base legal para el tratamiento de datos.*
- ***Fiabilidad:*** *Alucinaciones, tecnologías invasivas que introducen nuevos riesgos a los derechos y libertades de personas (Corrección, Portabilidad).*
- ***Medidas de Seguridad:*** *No siempre podemos controlar a donde fluye la información, o las medidas técnicas de seguridad que se implementan (T&Cs estándares).*
- ***Violaciones de datos personales:*** *Mientras más datos disponibles, más riesgos de violación.*

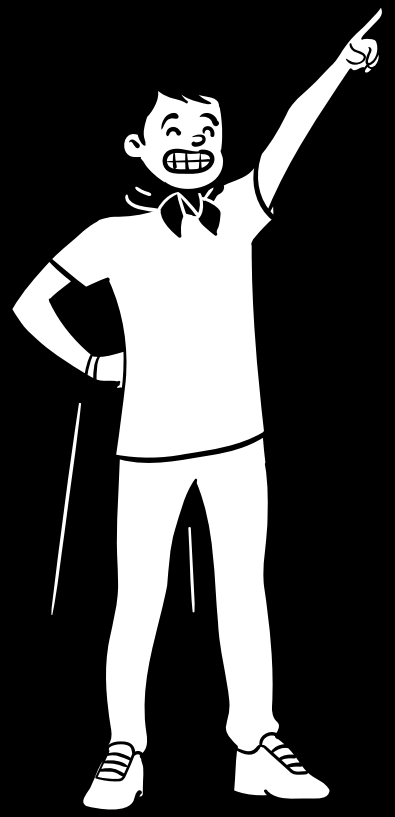
Combinado con

1) La falta de personal privacidad técnico
2) La rapidez de la inversión en IA.

QUÉ PODEMOS HACER?

Recomendación a profesionales de la Seguridad y la Privacidad.

1. No bloquear la innovación.
2. Convertirnos en “compañeros de combate” que apoya decisiones basadas en riesgos(derechos y libertades) y seguridad de la información.
3. Facilita un marco de apoyo para evaluar technologías de IA; (Adapte su PIA)



EJEMPLO CUESTIONARIO DE EVALUCACIÓN PARA IA

Pregunta	Riesgo Sí o No
1. Confidencialidad: ¿Se facilitará alguna información confidencial de nuestra empresa (información restringida o interna)?	
2. Privacidad: Se recopilará y procesará datos personales de Clientes o empleados ?¿Qué categoría de datos?	
3. Base legal o consentimiento: ¿Existe alguna base legal o consentimiento para tratar los datos personales en cuestión? ¿Esto se puede documentar? ¿Estamos siendo lo suficientemente transparente con ello?	
4. Propiedad y utilización de datos: ¿Quién tiene el derecho a la Propiedad Intelectual sobre los datos?¿Serán los datos accedidos/gestionados únicamente por el usuario dueño (y las partes interesadas)? ¿O se le otorgará al proveedor de IA el derecho de acceder, compartir, publicar, vender , etc estos datos como mejor le parezca? (Esto se puede verificar en los T&C del proveedor del producto de IA).	
5. Alojamiento de datos: ¿Residirán los datos en la Unión Europea o fuera? ¿Se alojarán internamente o con el proveedor de productos de IA?	
6. Madurez en cumplimiento: ¿Tiene el proveedor algún estándar internacional (ISO27001, SOC II, etc.) ¿Sigue las pautas del RGPD?	
7. Derechos de las personas físicas: ¿pueden los usuarios ejercer sus derechos y libertades de datos? (por ejemplo, solicitud de eliminación de datos o rectificación de datos)	
8. Transparencia ¿Entiende cómo funciona la solución de IA y podría explicar cómo funciona a las partes interesadas?	
9. Veracidad y exactitud ¿La solución de IA brinda resultados lo suficientemente confiables (precisos, imparciales, fácticos, etc.), para ser usado en la toma de decisiones empresariales?	
10. Medidas técnicas: ¿Conoce qué medidas técnicas facilita la solución de IA para proteger la información sometida? Ej. Cifrado de datos, anonimato de datos.	