EvenBetter

Skills curadas para agentes: apoyo inteligente y sensible al contexto en UX y accesibilidad de iOS

Marcello Gonzatto Birkan[1] | Daniela Brazolin Flauto[1] | Valéria Farinazzo Martins[2]

[1] Faculdade de Computação e Informática · [2] Programa de Pós-Graduação em Computação Aplicada
Universidade Presbiteriana Mackenzie · São Paulo, Brasil

LIVE · EN VIVO · AO VIVO

Follow the talk live, in your language

Sigue la charla en vivo, en tu idioma

Acompanhe a palestra ao vivo, no seu idioma

evenbetter-slides.mellofattura.com/live

FUERZA 01

accesibilidad digital

las guías existen, pero no siempre se vuelven requisitos

FUERZA 01

accesibilidad digital

las guías existen, pero no siempre se vuelven requisitos

FUERZA 02

IA en el desarrollo

los agentes ya trabajan dentro del flujo y llenan los vacíos

FUERZA 01

accesibilidad digital

las guías existen, pero no siempre se vuelven requisitos

CONFLICTO

investigación

FUERZA 02

IA en el desarrollo

los agentes ya trabajan dentro del flujo y llenan los vacíos

Acto I

El problema doble

La accesibilidad no se vuelve requisito. La IA llena los vacíos.

PROBLEMA

Una pantalla puede compilar y aun así dejar fuera a personas

VoiceOver

ESCUCHA LA PANTALLA



Texto más grande

LEE CON LETRA MÁS GRANDE



Control por voz

HABLA EN VEZ DE TOCAR



PROBLEMA

Una pantalla puede compilar y aun así dejar fuera a personas

VoiceOver

ESCUCHA LA PANTALLA



Solo oye "botón" e "imagen"

Texto más grande

LEE CON LETRA MÁS GRANDE



El texto queda cortado

Control por voz

HABLA EN VEZ DE TOCAR



Nada en la pantalla se llama Eliminar

Una pantalla puede compilar y aun así dejar fuera a personas



2,5

mil millones de personas necesitan productos de apoyo

✦ MÁS ALLÁ DEL ARTÍCULO — WHO & UNICEF · GLOBAL REPORT ON ASSISTIVE TECHNOLOGY, 2022

PROBLEMA

El 95,9% del millón de páginas de inicio más visitadas tiene fallas WCAG detectables



WEBAIM MILLION, 2026

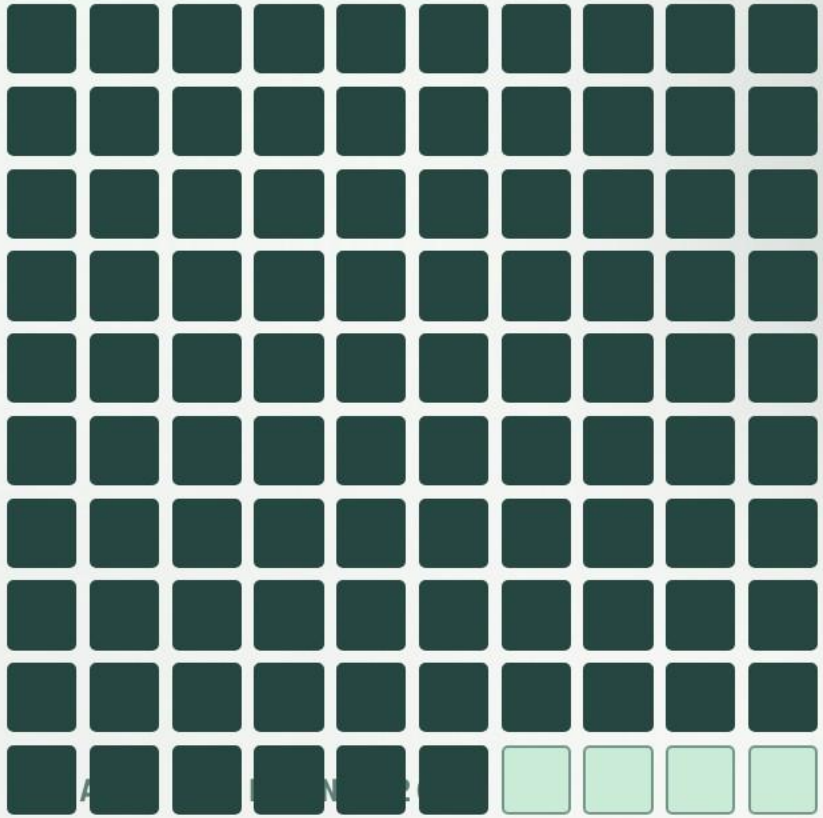
fallas detectadas

pasaron las pruebas automáticas

1 CUADRADO = 10.000 PÁGINAS DE INICIO

PROBLEMA

El 95,9% del millón de páginas de inicio más visitadas tiene fallas WCAG detectables



95,9%

DEL MILLÓN DE PÁGINAS DE INICIO MÁS VISITADAS

fallas detectadas

pasaron las pruebas automáticas

1 CUADRADO = 10.000 PÁGINAS DE INICIO

Sintáctico vs semántico

Las herramientas de hoy detectan solo parte del problema: no pueden juzgar el contexto ni la intención.

SINTÁCTICO

Falta el atributo, código mal formado.

```
Image("cart")
```

Sintáctico vs semántico

Las herramientas de hoy detectan solo parte del problema: no pueden juzgar el contexto ni la intención.

SINTÁCTICO

Falta el atributo, código mal formado.

```
Image("cart")
```

SEMÁNTICO

Atributo correcto, contenido equivocado.

```
Image("cart")
  .accessibilityLabel("image")
```

Sintáctico vs semántico

Las herramientas de hoy detectan solo parte del problema: no pueden juzgar el contexto ni la intención.

SINTÁCTICO

Falta el atributo, código mal formado.

```
Image("cart")
```

✓ EL VERIFICADOR LO MARCA

SEMÁNTICO

Atributo correcto, contenido equivocado.

```
Image("cart")
  .accessibilityLabel("image")
```

! PASA · VOICEOVER DICE “IMAGE”

Una encuesta a más de 49.000 desarrolladores de 177 países:

84%

usan IA en el trabajo o planean empezar.

STACK OVERFLOW DEVELOPER SURVEY, 2025

Una encuesta a más de 49.000 desarrolladores de 177 países:

84%

usan IA en el trabajo o planean empezar.

CONFIANZA EN LA PRECISIÓN DE LA IA

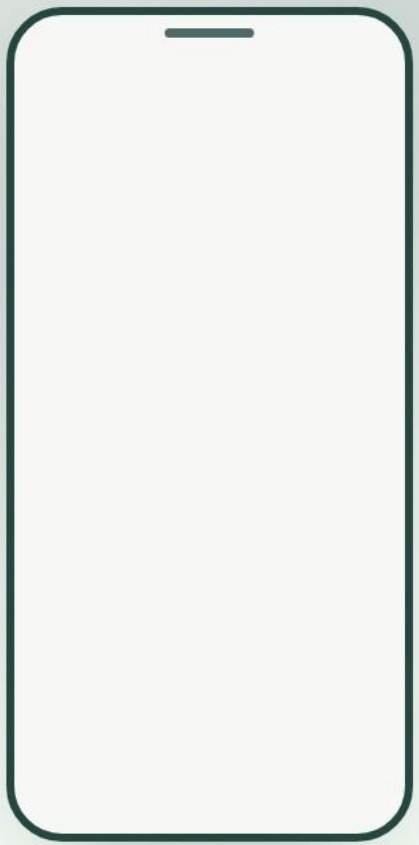


STACK OVERFLOW DEVELOPER SURVEY, 2025

PROBLEMA

Un prompt, muchas decisiones que nadie tomó

Crea una pantalla para listar y |



BARRERA DE ARTICULACIÓN · NIELSEN, 2023

PROBLEMA

Un prompt, muchas decisiones que nadie tomó

Crea una pantalla para listar y eliminar mis tareas.

Listo. Compila ✓



BARRERA DE ARTICULACIÓN · NIELSEN, 2023

PROBLEMA

Un prompt, muchas decisiones que nadie tomó

> Crea una pantalla para listar y eliminar mis tareas.

LO QUE EL DESARROLLADOR TENDRÍA QUE EXPLICITAR

01 **Plantear bien el problema**

02 **Dar contexto y restricciones**

03 **Definir criterios de calidad**

04 **Especificar el resultado esperado**



•OBJETIVO DE LA INVESTIGACIÓN

Mejorar la *accesibilidad digital* y abrir el *uso avanzado* de IA en el desarrollo de apps iOS

•OBJETIVO DE LA INVESTIGACIÓN

Mejorar la *accesibilidad digital* y abrir el *uso avanzado* de IA en el desarrollo de apps iOS

RQ1

¿Cómo pueden las skills curadas convertir las guías de UX y accesibilidad en procedimiento?

→ EVENBETTER, EL ARTEFACTO

RQ2

¿Cómo perciben y usan los desarrolladores iOS un flujo de trabajo con skills?

→ UN ESTUDIO CON 10 DESARROLLADORES

Acto II

De la guía al procedimiento

Las skills curadas convierten las
guías en pasos que el agente sigue.

DOS MOMENTOS

> Podemos usar **IA** en dos momentos de la *accesibilidad* de una app: antes de que exista la pantalla y después de publicarla.

01 DESDE EL INICIO

Planificación

accesibilidad desde el diseño

SCREENAUDIT · ZHONG ET AL., CHI 2025

DOS MOMENTOS

69,2%
AUDITORÍA CON LLM
VERIFICADOR
31,3%

Podemos usar **IA** en dos momentos de la *accesibilidad* de una app: antes de que exista la pantalla y después de publicarla.

01 DESDE EL INICIO

Planificación

accesibilidad desde el diseño

MENOS EXPLORADO

02 YA EN EL PRODUCTO

Detectar y corregir

AQUÍ ACTÚA LA MAYORÍA DE LAS HERRAMIENTAS

- ScreenAudit
- AccessGuru
- FixAlly
- DesignRepair

SCREENAUDIT · ZHONG ET AL., CHI 2025

DOS MOMENTOS

69,2%
AUDITORÍA CON LLM
VERIFICADOR
31,3%

Podemos usar **IA** en dos momentos de la *accesibilidad* de una app: antes de que exista la pantalla y después de publicarla.

01 DESDE EL INICIO

Planificación

accesibilidad desde el diseño

MENOS EXPLORADO

02 YA EN EL PRODUCTO

Detectar y corregir

AQUÍ ACTÚA LA MAYORÍA DE LAS HERRAMIENTAS

ScreenAudit · AccessGuru · FixAlly · DesignRepair

• **EvenBetter** CUBRE LOS DOS MOMENTOS

SCREENAUDIT · ZHONG ET AL., CHI 2025

Un pedido corto carga todo el procedimiento

```
$ /evenbetter-analyze
```

AGENT SKILLS SPECIFICATION, 2025

Un pedido corto carga todo el procedimiento

```
$ /evenbetter-analyze

o

"revisa mi app según la
HIG de Apple"
```

EVENBETTER-ANALYZE/SKILL.MD

description: auditar un proyecto iOS o SwiftUI	Cuándo se usa
preguntar solo si el alcance no está claro	Qué preguntar
leer primero el corpus: 49 cláusulas	Qué leer
1 escanear · 2 verificar · 3 informar	Pasos
informe HTML con archivo y línea	Salida

CURADA ANTES DE QUE EL AGENTE CORRA

AGENT SKILLS SPECIFICATION, 2025

EVIDENCIA

Importa quién escribe la skill

CURADA

Escrita desde fuentes verificables · revisada · versionada

AUTOGENERADA

Escrita por el agente en el momento · propensa a alucinaciones

SKILLSBENCH, 2026 · TASA MEDIA DE APROBACIÓN FRENTE A UNA LÍNEA BASE SIN SKILLS

EVIDENCIA

Importa quién escribe la skill

CURADA

Escrita desde fuentes verificables · revisada · versionada

AUTOGENERADA

Escrita por el agente en el momento · propensa a alucinaciones



SKILLSBENCH, 2026 · TASA MEDIA DE APROBACIÓN FRENTE A UNA LÍNEA BASE SIN SKILLS

POSTURA

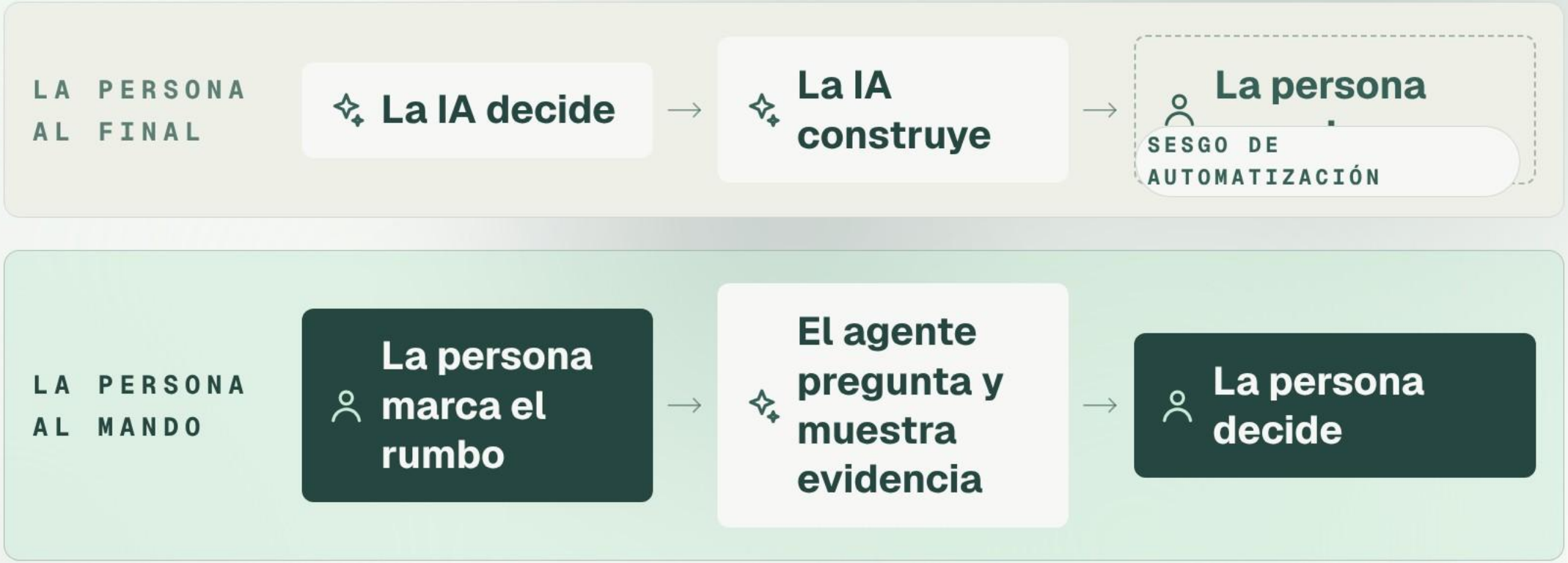
La persona al mando, no al final



SHNEIDERMAN, 2020 · PARASURAMAN & MANZEY, 2010

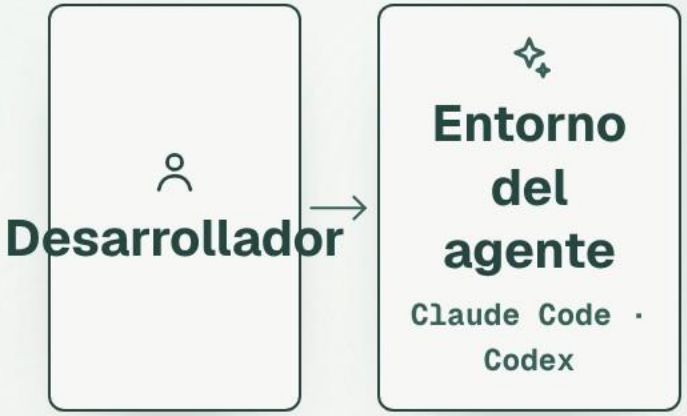
POSTURA

La persona al mando, no al final

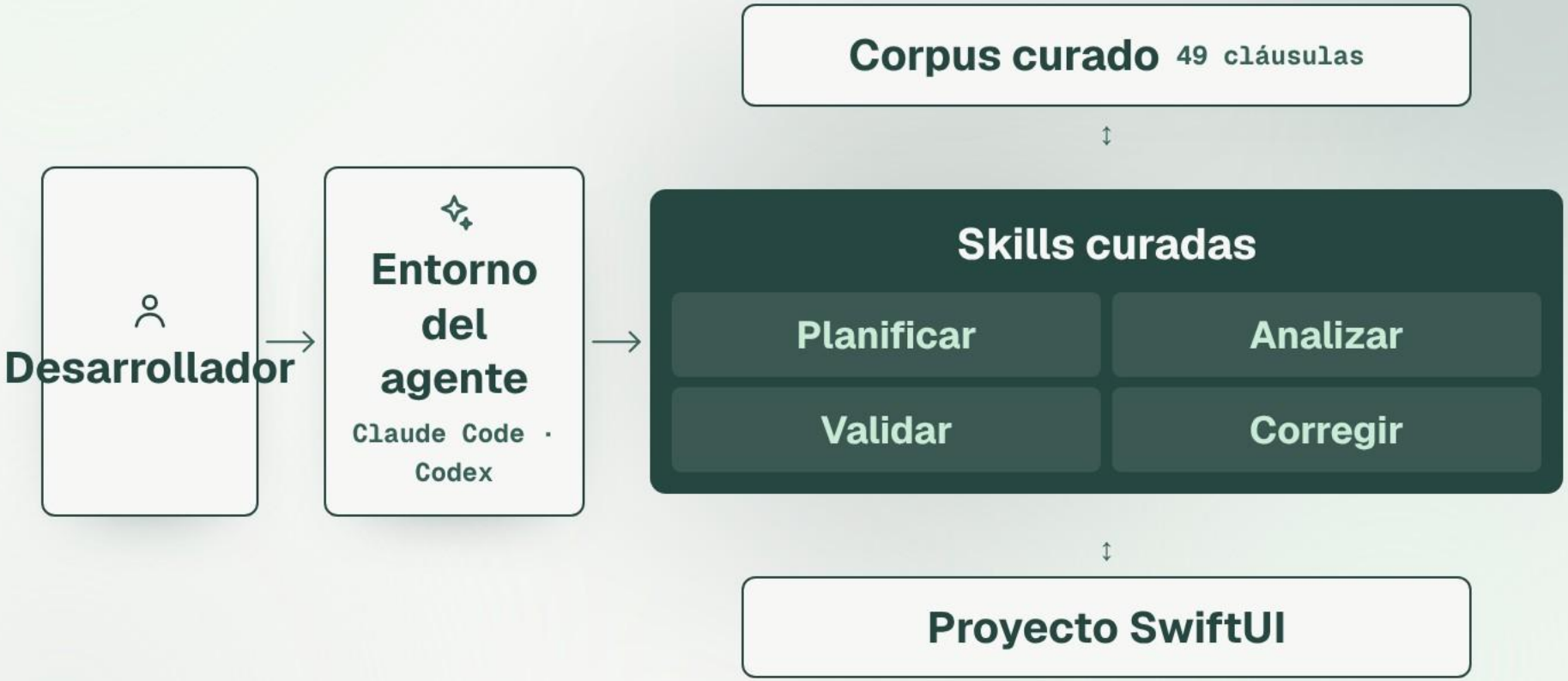


SHNEIDERMAN, 2020 · PARASURAMAN & MANZEY, 2010

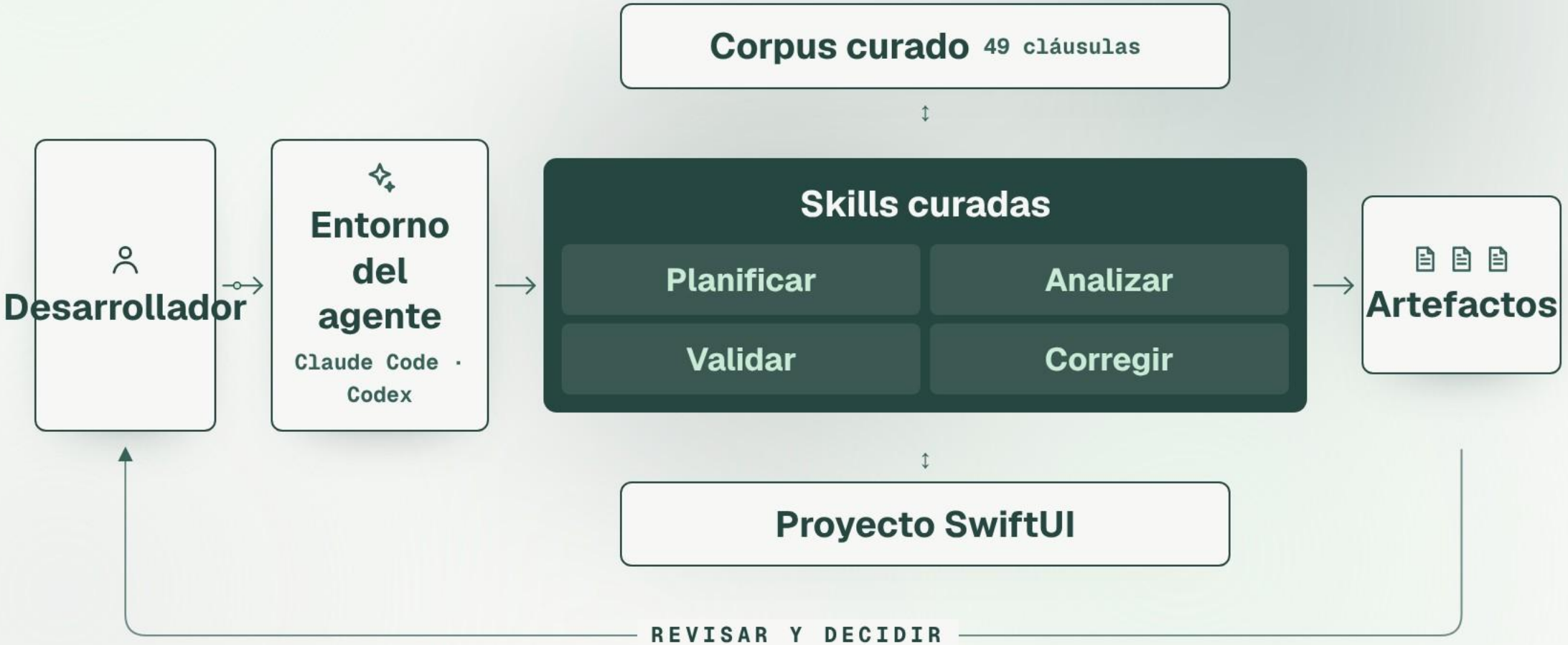
Cómo funciona el sistema



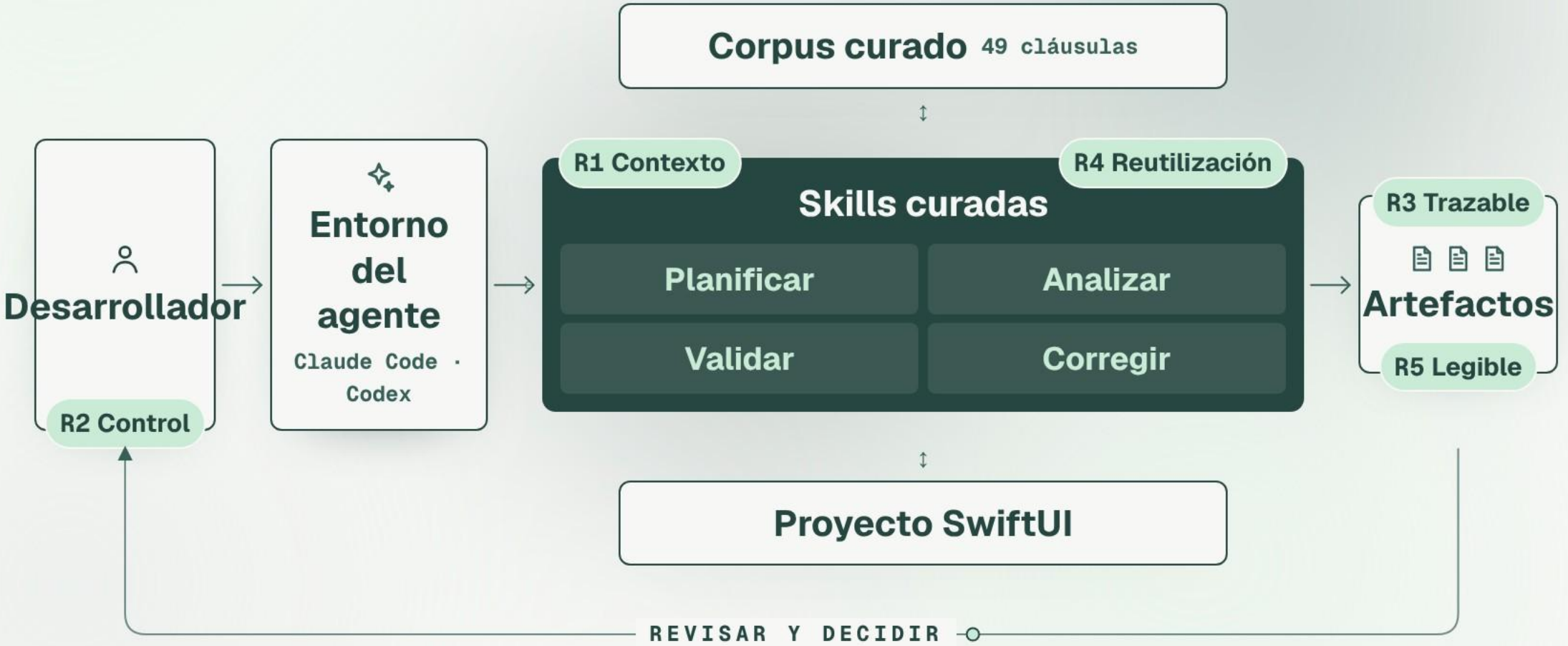
Cómo funciona el sistema



Cómo funciona el sistema



Cómo funciona el sistema



CORPUS CURADO

49 cláusulas, y cada hallazgo cita una



Cada cláusula tiene un ID, una fuente pública, una verificación y un patrón correcto.

APUNTA A UN PUNTO PARA LEER SU CLÁUSULA · TOCA PARA FIJARLA

ESCRITAS A PARTIR DE APPLE HIG · APPLE ACCESSIBILITY · WCAG 2.2

CORPUS CURADO

49 cláusulas, y cada hallazgo cita una



CORPUS/IOS/TYPOGRAPHY.MD

```
## TYPO-UI-001 - Prefer system text styles

Source: Apple HIG: Typography

Check. Flag fixed .font(.system(size:)) on semantic text

Text("Account").font(.title2)
```

ESCRITAS A PARTIR DE APPLE HIG · APPLE ACCESSIBILITY · WCAG 2.2

CORPUS CURADO

49 cláusulas, y cada hallazgo cita una



9
8
8
7
7
6
4

CORPUS/IOS/TYPOGRAPHY.MD

ID estable

TYPO-UI-001 - Prefer system text styles

Fuente pública

Source: Apple HIG: Typography

Verificación concreta

Check. Flag fixed `.font(.system(size:))` on semantic text

Patrón correcto

```
Text("Account").font(.title2)
```

→ EL HALLAZGO DE LA FIG. 2 CITA ESTA CLÁUSULA

ESCRITAS A PARTIR DE APPLE HIG · APPLE ACCESSIBILITY · WCAG 2.2

Acto III

La prueba en uso

Dos skills en acción y luego diez desarrolladores usándolas.

Primero planifica.

En una épica, pregunta antes de escribir código.

✓ opciones cerradas
· documentación de Apple consultada

PLAN

PÚBLICO
Gente que sigue series de TV.

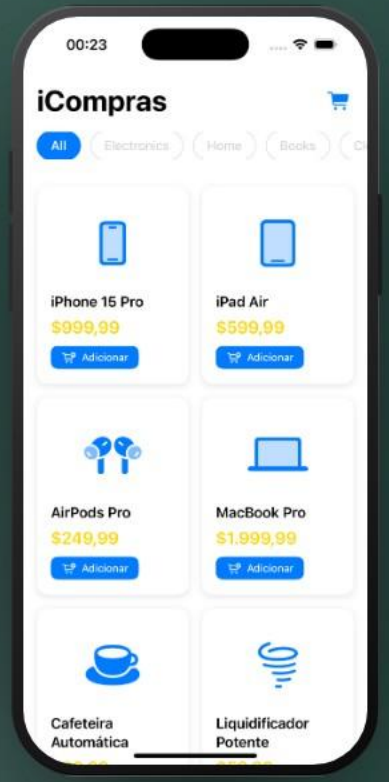
PANTALLAS
TabView · 3 pestañas.

PERSISTENCIA
SwiftData · un usuario.

ACCESIBILIDAD
Se adapta. Sin truncar.

→ PLAN + TAREAS DE DESARROLLO

La skill que *audita* una app que ya existe.



APP DE DEMO · ICOMPRAS

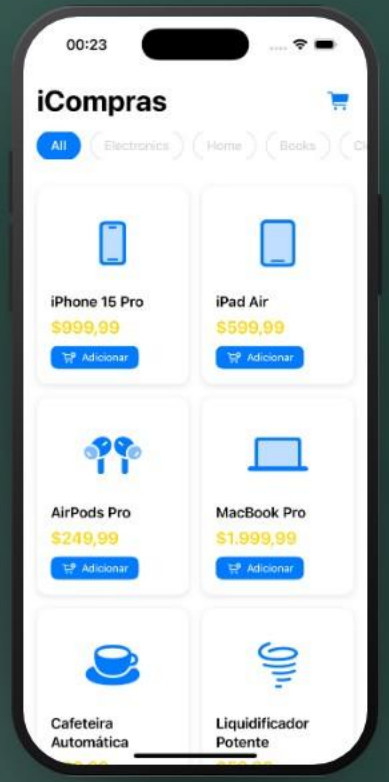
PASO DE LA DEMO

Escribe la skill en Claude Code

```
$ /evenbetter-analyze
```

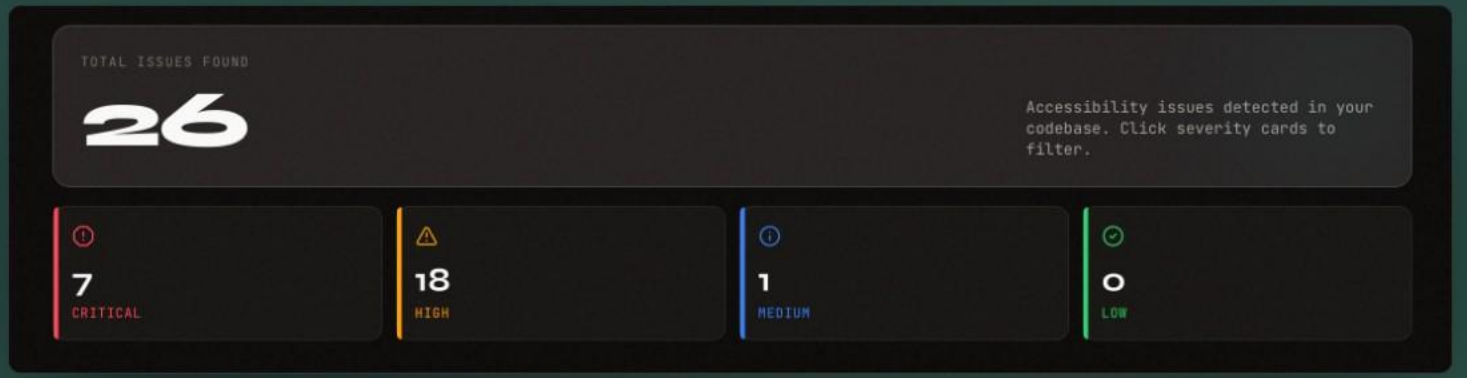
ENTRADA → ANÁLISIS DEL PROYECTO → CORPUS EVENBETTER → INFORME

La skill que *audita* una app que ya existe.



APP DE DEMO · ICOMPRAS

EL INFORME QUE ESCRIBIÓ LA SKILL



ENTRADA → ANÁLISIS DEL PROYECTO → CORPUS EVENBETTER → INFORME

Un hallazgo, de la evidencia a la corrección

En el estudio, la auditoría de Aroma Care encontró 55 problemas: 15 críticos y 40 altos.

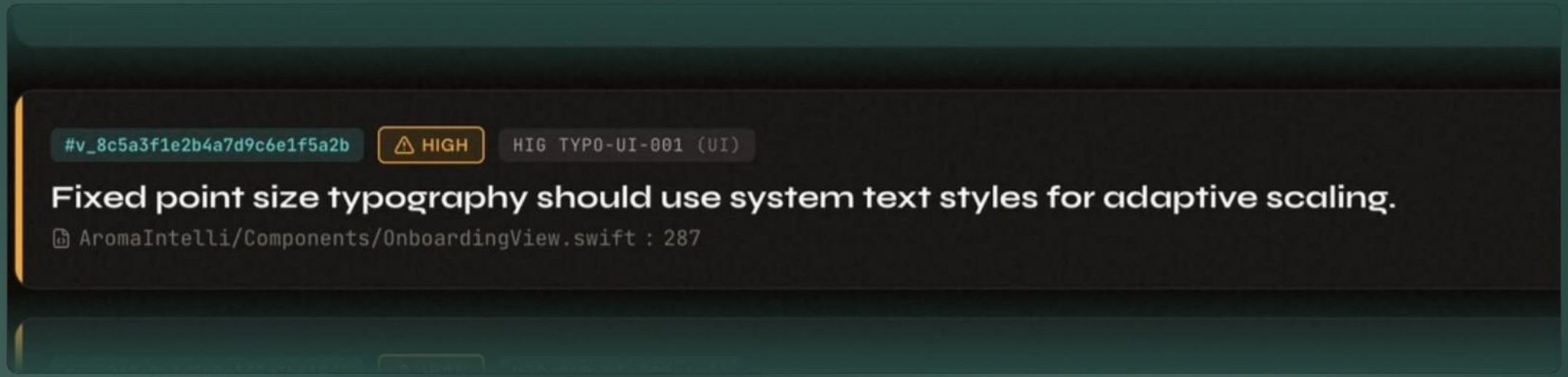


FIG. 2 · PRIMER HALLAZGO · CITA TYPO-UI-001

Un hallazgo, de la evidencia a la corrección

En el estudio, la auditoría de Aroma Care encontró 55 problemas: 15 críticos y 40 altos.

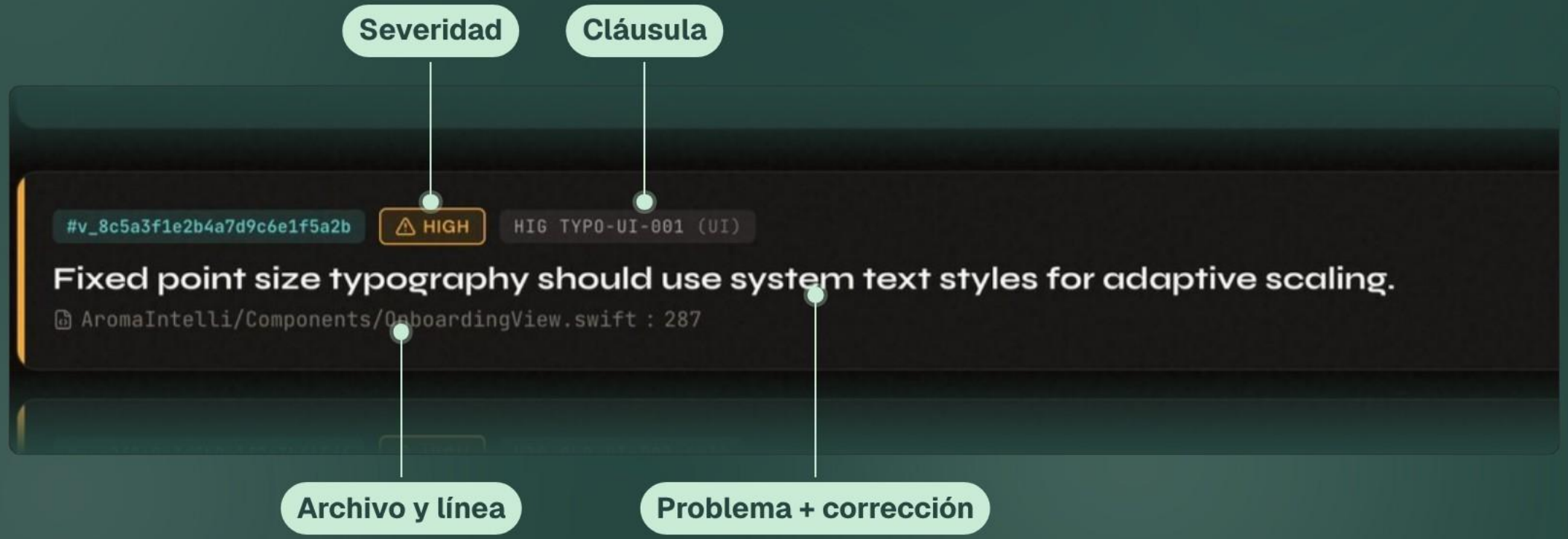


FIG. 2 · PRIMER HALLAZGO · CITA TYPO-UI-001

Diez desarrolladores iOS, una sesión de 30 minutos

EXPERIENCIA EN IOS



< 1 año | 1–2 años | 3–5 años

PENSAMIENTO EN VOZ ALTA · APROBADO POR EL COMITÉ DE ÉTICA EN INVESTIGACIÓN DE MACKENZIE

ESTUDIO

Diez desarrolladores iOS, una sesión de 30 minutos

EXPERIENCIA EN IOS



AGENTE USADO



PENSAMIENTO EN VOZ ALTA · APROBADO POR EL COMITÉ DE ÉTICA EN INVESTIGACIÓN DE MACKENZIE

ESTUDIO

Diez desarrolladores iOS, una sesión de 30 minutos

EXPERIENCIA EN IOS



< 1 año · 1–2 años · 3–5 años

AGENTE USADO



Claude Code · Copilot · Ninguno

≈ 30 MIN

Perfil · **Sitio** · Planificar · **Auditar** · SUS

PENSAMIENTO EN VOZ ALTA · APROBADO POR EL COMITÉ DE ÉTICA EN INVESTIGACIÓN DE MACKENZIE

RESULTADOS

SUS medio de 79,5, por encima de la referencia de 68

79,5

SUS MEDIO · N = 10

Referencia de la industria: 68

De bueno a excelente

FIG. 3 · SUS, BROOKE 1996 · BANDAS DE BANGOR ET AL., 2009

SUS medio de 79,5, por encima de la referencia de 68

79,5

SUS MEDIO · N = 10

Mediana	81,25
DE	15,1
Rango	57,5–100



FIG. 3 · SUS, BROOKE 1996 · BANDAS DE BANGOR ET AL., 2009

SUS medio de 79,5, por encima de la referencia de 68

79,5

SUS MEDIO · N = 10

Mediana	81,25
DE	15,1
Rango	57,5–100

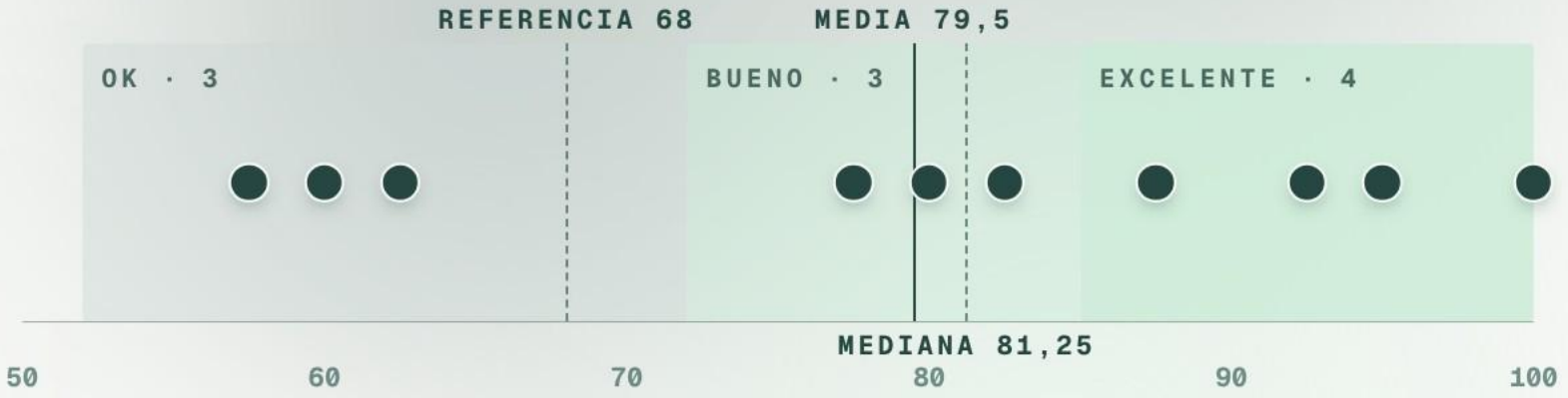


FIG. 3 · SUS, BROOKE 1996 · BANDAS DE BANGOR ET AL., 2009

RESULTADOS

Los diez participantes leyeron bien el informe

01 · DESCUBRIR

4/10



usaron el chat del sitio

15 CRÍTICOS + 40 ALTOS · TODOS LOS CONTARON BIEN

RESULTADOS

Los diez participantes leyeron bien el informe

01 · DESCUBRIR

4/10



usaron el chat del sitio

02 · PLANIFICAR

4/10



llamaron a la skill epic

15 CRÍTICOS + 40 ALTOS · TODOS LOS CONTARON BIEN

RESULTADOS

Los diez participantes leyeron bien el informe

01 · DESCUBRIR

4/10



usaron el chat del sitio

02 · PLANIFICAR

4/10



llamaron a la skill epic

03 · AUDITAR

8/10



analizar → validar

15 CRÍTICOS + 40 ALTOS · TODOS LOS CONTARON BIEN

RESULTADOS

Los diez participantes leyeron bien el informe

01 · DESCUBRIR

4/10



usaron el chat del sitio

02 · PLANIFICAR

4/10



llamaron a la skill epic

03 · AUDITAR

8/10



analizar → validar

04 · LEER

10/10



encontraron los 55 hallazgos

15 CRÍTICOS + 40 ALTOS · TODOS LOS CONTARON BIEN

En sus propias palabras

Respuestas abiertas del postest · traducidas del portugués

> “Me gustó mucho cómo está organizado el informe de errores y cómo se presenta toda la información. Creo que también es una gran forma de aprender.”
>
> *“Gostei muito da forma como o relatório de erros está organizado e todas as informações são apresentadas. Acredito que também é uma ótima forma de aprendizado.”*

CUESTIONARIO POSTEST · N = 10

En sus propias palabras

Respuestas abiertas del postest · traducidas del portugués

> “Me gustó mucho cómo está organizado el informe de errores y cómo se presenta toda la información. Creo que también es una gran forma de aprender.”
>
> *“Gostei muito da forma como o relatório de erros está organizado e todas as informações são apresentadas. Acredito que também é uma ótima forma de aprendizado.”*

> “Las dificultades que tuve vinieron de mi poca experiencia con IA, pero el sitio fue un gran apoyo para mis dudas.”
>
> *“As dificuldades encontradas foram relacionadas a inexperiência com o uso de IA, mas o site foi um ótimo suporte para as dúvidas.”*

CUESTIONARIO POSTEST · N = 10

Epílogo

Lo que queda

Qué cambia para las herramientas y qué sigue abierto.

DISCUSIÓN

Qué significa esto para las herramientas con agentes

01

Un prompt inicia una interacción

Qué significa esto para las herramientas con agentes

01

Un prompt inicia una interacción

02

Mostrar evidencia, no veredictos

DISCUSIÓN

Qué significa esto para las herramientas con agentes



01

Un prompt inicia una interacción



02

Mostrar evidencia, no veredictos



03

Preguntar cuando la intención decide

DISCUSIÓN

Qué significa esto para las herramientas con agentes



01

Un prompt inicia una interacción



02

Mostrar evidencia, no veredictos



03

Preguntar cuando la intención decide



04

El onboarding es parte del sistema

Con las skills, agregué las siete a mi propia app del App Store



Weightics

Weight Trend & BMI

APP STORE

Accesibilidad	7 / 7
VoiceOver	✓
Control por voz	✓
Texto más grande	✓
Interfaz oscura	✓
Diferenciar sin usar solo el color	✓
Contraste suficiente	✓
Movimiento reducido	✓

APP STORE · WEIGHTICS: WEIGHT TREND & BMI · NOV. 2026

Con las skills, agregué las siete a mi propia app del App Store



Weightics

Weight Trend & BMI

APP STORE

Accesibilidad	7 / 7
VoiceOver	✓ — A11Y-UI-001
Control por voz	✓ — COMP-UI-001
Texto más grande	✓ — TYPO-A11Y-003
Interfaz oscura	✓ — CLR-UI-003
Diferenciar sin usar solo el color	✓ — CLR-UX-001
Contraste suficiente	✓ — CLR-A11Y-001
Movimiento reducido	✓ — A11Y-UX-001

Cada función corresponde a una cláusula del corpus

APP STORE · WEIGHTICS: WEIGHT TREND & BMI · NOV. 2026

LIMITACIONES

Lo que este estudio todavía no muestra

LIMITACIÓN

- Sin línea base
- Precisión no medida
- 10 devs, 1 app
- Solo iOS

LIMITACIONES

Lo que este estudio todavía no muestra

LIMITACIÓN	LO QUE SIGUE
Sin línea base	Comparar sin skills
Precisión no medida	Revisión experta de los hallazgos
10 devs, 1 app	Más apps y equipos
Solo iOS	Android, web y juegos

•CONCLUSIÓN

El conocimiento curado guía al agente; *la decisión sigue siendo humana.*

• CONCLUSIÓN

El conocimiento curado guía al agente; *la decisión sigue siendo humana.*

17	49	79,5	10/10
skills instalables	cláusulas de iOS	SUS medio	leyeron bien el informe

QUESTIONS · PREGUNTAS · PERGUNTAS

Thank you.

Gracias · Obrigado

Marcello Gonzatto Birkan · Daniela Brazolin Flauto · Valéria Farinazzo Martins

Universidade Presbiteriana Mackenzie · mgbirkan@icloud.com

evenbetter.mellofattura.com