



Índice de Contenidos

Introducción y Propósito

¿Qué es la Inteligencia Artificial para IRIA?

¿Cómo Utilizan los Clientes los Resultados?

¿Por Qué Utilizamos IA?

Arquitectura Basada en el Control

Experiencia del Usuario y Configuración

Proveedores de IA y Fuentes de Datos

Explicabilidad y Limitaciones

Pruebas y Validación del Sistema

Mantenimiento y Actualización

Partes Interesadas y Gestión de Riesgos

Apéndices

1. INTRODUCCIÓN Y PROPÓSITO

El presente documento constituye la Declaración de Explicabilidad de **IRIA**. Su propósito es proporcionar a nuestros clientes, candidatos y demás partes interesadas una visión transparente y rigurosa sobre cómo se diseña, implementa y gobierna la tecnología de Inteligencia Artificial (IA) en nuestra plataforma de selección de talento.

IRIA considera el desarrollo ético de la IA, la seguridad de los datos y la privacidad como valores corporativos fundamentales. Este documento es una manifestación de dicho compromiso y tiene como objetivo asistir a nuestros clientes en el cumplimiento de sus obligaciones como

"Responsables del Tratamiento" de datos, en conformidad con normativas como el Reglamento General de Protección de Datos (RGPD) de la UE.

Documento Vivo: Este es un documento que será actualizado periódicamente para reflejar la evolución de nuestros sistemas, procesos y el Estado del Arte en gobernanza de IA.

2. ¿QUÉ ES LA INTELIGENCIA ARTIFICIAL PARA IRIA?

Para nosotros, la IA es un conjunto de tecnologías avanzadas, principalmente del campo del **Procesamiento del Lenguaje Natural (NLP)**, que permiten a las máquinas entender, procesar y generar lenguaje humano. En **IRIA**, esto se materializa en un sistema capaz de mantener una

conversación coherente, interpretar las respuestas de un candidato y evaluarlas de manera objetiva y estructurada contra un conjunto de criterios predefinidos.

3. ¿CÓMO UTILIZAN LOS CLIENTES LOS RESULTADOS DE IRIA?

IRIA está diseñada para integrarse de forma fluida en el flujo de trabajo de un reclutador moderno, automatizando determinadas fases del proceso para que los equipos de selección puedan centrarse en la toma de decisiones estratégicas.

El proceso típico es el siguiente:

- 1. Configuración del Puesto:** El reclutador crea una nueva posición en la plataforma, definiendo no solo los requisitos técnicos y de experiencia, sino también la cultura y filosofía de la empresa.
- 2. Gestión de Candidatos:** El reclutador añade candidatos a la plataforma y los asigna a una posición.
- 3. Convocatoria a Entrevista:** El reclutador genera las credenciales de los candidatos y un enlace para que puedan agendar la entrevista.

4. **Realización de la Entrevista:** El candidato realiza la entrevista con el avatar de IRIA. Las entrevistas son grabadas para su posterior revisión.
5. **Análisis y Decisión:** Una vez completada la entrevista y su posterior análisis por la IA, el reclutador puede ver un listado de candidatos con sus puntuaciones globales.
6. **Toma de Decisión Humana:** Con toda esta información estructurada y objetiva, el reclutador toma la decisión final sobre qué candidatos avanzan a las siguientes fases del proceso.

Importante: La empresa que utiliza IRIA actúa siempre como la "controladora de datos" bajo el GDPR, mientras que IRIA es el "procesador de datos". Cualquier solicitud de un candidato para cuestionar una decisión debe dirigirse a la empresa contratante.

4. ¿POR QUÉ UTILIZAMOS IA?

El proceso de selección de personal tradicional, aunque valioso, se enfrenta a desafíos significativos. La revisión manual de cientos de currículums, la realización de entrevistas, la redacción de informes técnicos y la coordinación de agendas consumen una enorme cantidad de tiempo y recursos.

IRIA utiliza IA para abordar directamente estos problemas, ofreciendo un proceso más **justo, consistente y eficiente**.

Ventajas de la IA de IRIA

- **Reducción y Mitigación de Sesgos:** A diferencia de los humanos, la IA evalúa a cada candidato aplicando exactamente los mismos criterios predefinidos. Nuestra validación empírica con 2,498 candidatos sintéticos ha confirmado la ausencia completa de sesgos discriminatorios, con todos los effect sizes por debajo de 0.1.

- **Consistencia en la Evaluación:** IRIA garantiza que cada entrevista se realice y evalúe de la misma manera. Esta consistencia ha sido validada empíricamente con una puntuación del 91.4% en consistencia inter-variacional y 97.9% en fiabilidad test-retest.
- **Igualdad de Oportunidades:** Al ser una plataforma online y automatizada, IRIA elimina barreras geográficas y temporales.
- **Análisis Profundo y Perspicaz:** IRIA está diseñada para ser un entrevistador perspicaz que pide ejemplos concretos y profundiza en la veracidad de las competencias.
- **Mejores Decisiones Basadas en Datos:** IRIA proporciona informes detallados y métricas objetivas para comparar candidatos.
- **Eficiencia y Ahorro de Costes:** Al automatizar las entrevistas y la redacción de informes, IRIA libera al equipo de Recursos Humanos para tareas de mayor valor.

5. ¿CÓMO DISEÑAMOS NUESTRA IA? UNA ARQUITECTURA BASADA EN EL CONTROL

La arquitectura de IA de IRIA se fundamenta en una filosofía clave: **guiar y controlar modelos de vanguardia, en lugar de entrenar modelos limitados con datos del pasado.**

Nuestro enfoque es deliberadamente diferente y representa la próxima generación de aplicación de IA. En lugar de construir un modelo "pequeño" y estático, aprovechamos el poder y el conocimiento masivo y actualizado de los **Modelos de Lenguaje Avanzados (LLMs) de Google (Gemini)**.

El Marco de Evaluación por Competencias Estructuradas (MEC)

El MEC es un sofisticado sistema de gobernanza y control que actúa como una "plantilla computacional" sobre el LLM. En lugar de hacer una pregunta abierta a la IA, forzamos al modelo a razonar dentro de los límites de nuestro marco. El MEC descompone la evaluación en un análisis multidimensional riguroso, obligando a la IA a justificar cada puntuación basándose en evidencia textual extraída de la entrevista y el CV.

Concretamente, nuestro marco actual evalúa **hasta 64 sub-métricas específicas** para componer las competencias de alto nivel, garantizando granularidad y solidez en el análisis.

Ventajas Estratégicas del MEC

- **Dinamismo y Adaptabilidad:** El MEC se adapta a cualquier puesto, categoría profesional o cultura de empresa en tiempo real.
- **Precisión Contextual:** Al usar un LLM de propósito general como Gemini, nuestro sistema posee una comprensión mucho más rica del lenguaje y los contextos técnicos.
- **Seguridad y Ética Amplificadas:** Nos beneficiamos de las inversiones de Google en seguridad y mitigación de sesgos, añadiendo nuestra propia capa de control.

Flujo de Análisis de Entrevista

1. **Recopilación de Contexto:** El sistema agrupa toda la información relevante: transcripción de la entrevista, texto del CV, contexto de la posición y cultura de la empresa.
2. **Aplicación del MEC sobre Gemini:** La información se procesa a través del modelo Gemini, pero la salida es estrictamente gobernada por nuestro MEC.
3. **Doble Control de Seguridad:** Implementamos un sistema de doble control para detectar y prevenir preguntas ilegales o discriminatorias.
4. **Validación y Almacenamiento:** La respuesta estructurada es validada por nuestro sistema antes de presentarse al reclutador.

6. EXPERIENCIA DEL USUARIO Y CONFIGURACIÓN DEL SISTEMA

Nuestro objetivo es que la experiencia con **IRIA** sea potente para los reclutadores y justa y sencilla para los candidatos. La plataforma está diseñada para ser un centro de mando integral para la adquisición de talento.

Formación del Cliente

Proporcionamos a nuestros clientes documentación completa sobre cómo interpretar los informes de análisis, el significado de cada métrica y las mejores prácticas para integrar **IRIA** en un proceso de selección equitativo.

Herramientas de Gestión del Reclutador

- **Panel de Control Centralizado:** Los reclutadores gestionan todo el ciclo de vida del candidato desde un único lugar.
- **Calendario Integrado:** Visualizan y gestionan las entrevistas programadas, pasadas y futuras.
- **Seguimiento de Onboarding:** Herramientas para hacer seguimiento del proceso de incorporación.

Configuración por el Cliente

- Definición de puestos y cultura de empresa
- Gestión de la comunicación con los candidatos
- Gestión de posiciones y toma de decisiones estratégica

Configuración Fija (Por Diseño)

Para garantizar la máxima consistencia y equidad, hay aspectos que no son configurables por el cliente:

- **Preguntas de la Entrevista:** Las preguntas son generadas adaptativamente por nuestro sistema para evaluar de forma estandarizada las competencias.
- **Ponderación Dinámica y Contextual de Competencias:** El MEC instruye al modelo de IA para realizar una evaluación integral que pondera dinámicamente la relevancia de cada competencia.

7. PROVEEDORES DE IA Y FUENTES DE DATOS

En **IRIA**, creemos en la transparencia sobre las herramientas que utilizamos y el origen de los datos que impulsan nuestros análisis.

Proveedores Externos de Componentes de IA

- **Google Gemini (Modelo de Lenguaje Avanzado):** Utilizamos los potentes modelos de lenguaje de Google como el motor principal para nuestras capacidades de comprensión y generación de lenguaje natural.
- **HeyGen (Generador de Avatares y Voz):** La representación visual de nuestro entrevistador de IA y la síntesis de su voz son proporcionadas por la plataforma de HeyGen.
- **Google Cloud Speech-to-Text:** Para convertir las respuestas habladas de los candidatos en texto analizable, utilizamos la API Speech-to-Text de Google Cloud.

Fuentes de Datos para el Sistema

No entrenamos nuestros propios modelos de lenguaje desde cero. Nuestro enfoque se basa en guiar modelos fundacionales pre-entrenados. Por lo tanto, no utilizamos los datos de nuestros clientes para entrenar o re-entrenar los modelos base.

Los datos de las interacciones en la plataforma se utilizan de forma agregada y anonimizada únicamente con el fin de mejorar y validar la efectividad de nuestros procesos y de nuestro Marco de Evaluación por Competencias Estructuradas (MEC).

8. EXPLICABILIDAD DE LOS MODELOS Y SUS LIMITACIONES

Uno de nuestros principios fundamentales es la transparencia. Entendemos que las decisiones de contratación son críticas, y los usuarios de **IRIA** merecen saber cómo se generan nuestras evaluaciones.

¿Cómo se Explican las Puntuaciones?

La puntuación final de un candidato no es una "caja negra". Es el resultado directo del análisis del modelo de IA, rigurosamente delimitado por nuestro Marco de Evaluación por Competencias Estructuradas (MEC).

La lógica de la evaluación es inherentemente auditable. El sistema ejecuta una tarea específica, buscando evidencia textual de competencias específicas, calificándolas en escalas predefinidas y citando las frases que sustentan cada calificación. Este proceso se repite a través de **64 puntos de análisis distintos**.

Informes Detallados y Grabaciones

Nuestros informes no solo muestran una puntuación final, sino que desglosan las puntuaciones por cada competencia evaluada e incluyen las citas extraídas directamente de la

respuesta del candidato. Además, el reclutador siempre tiene acceso a la grabación completa de la entrevista.

Limitaciones

Es crucial reconocer las limitaciones de cualquier sistema de IA, incluido el nuestro:

- **La IA no es un Oráculo:** Nuestras evaluaciones son una herramienta de apoyo para la toma de decisiones, no un veredicto final.
- **Sesgos Inherentes en los LLMs:** Los modelos de lenguaje se entrenan con enormes cantidades de texto que inevitablemente contienen sesgos sociales y culturales.
- **Errores de Transcripción:** Ningún sistema de Speech-to-Text es perfecto. Puede haber errores especialmente con acentos fuertes o terminología técnica.
- **Interpretación del Lenguaje Figurado:** Los LLMs pueden tener dificultades para interpretar el sarcasmo, la ironía o el lenguaje muy figurado.

La decisión final de contratación **siempre** recae en el cliente. Recomendamos encarecidamente que se realice una revisión humana completa de la entrevista grabada ante cualquier duda.

9. PRUEBAS Y VALIDACIÓN DEL SISTEMA DE IA

La confianza en nuestro sistema se basa en un proceso continuo y riguroso de pruebas y validación que ha demostrado empíricamente resultados excepcionales. Nuestra más reciente validación exhaustiva, procesando **2,498 candidatos sintéticos**, ha confirmado una **consistencia inter-variacional del 91.4%**, una **fiabilidad test-retest del 97.9%**, y la **ausencia completa de sesgos discriminatorios** (todos los effect sizes < 0.1).

Testing Sintético de Análisis (TSA)

Como parte de nuestro compromiso con la excelencia en la validación y la mejora continua, hemos desarrollado un **Sistema de Testing Sintético de Análisis (TSA)** que constituye una metodología avanzada y rigurosa para la evaluación automatizada de sistemas de IA en selección de personal.

¿Qué es el TSA?

El TSA es un sistema automatizado que genera y ejecuta entrevistas sintéticas a gran escala para validar de manera científica la consistencia, equidad y calidad de nuestros análisis de IA.

Metodología Científica

El TSA genera automáticamente un volumen significativo de perfiles sintéticos derivados de perfiles base representativos, aplicando variaciones sistemáticas que reflejan la diversidad real de candidatos. Un aspecto fundamental es que las entrevistas

sintéticas se generan de forma completamente independiente de los criterios de evaluación posteriores, eliminando cualquier posibilidad de sesgo de confirmación.

Tipos de Validación Estadística

El TSA implementa múltiples metodologías de validación estadística para garantizar la robustez científica de nuestras evaluaciones:

- **Validación de Fiabilidad Test-Retest:** Medimos la consistencia temporal con repeticiones exactas del mismo perfil, alcanzando un 97.9% de fiabilidad.
- **Análisis de Equidad Demográfica:** Calculamos el tamaño del efecto (Cohen's d) para evaluar diferencias entre grupos demográficos.
- **Análisis de Consistencia Inter-variacional:** Evaluamos la estabilidad de puntuaciones ante variaciones demográficas irrelevantes, alcanzando un 91.4% de consistencia global.
- **Estadísticas Descriptivas Comprehensivas:** Calculamos media, desviación estándar, rangos, e intervalos de confianza del 95% para todas las métricas.
- **Análisis Multi-dimensional de Sesgos:** Evaluamos 10 dimensiones diferentes de posible sesgo.
- **Análisis de Variabilidad de Evaluaciones:** Las métricas de dispersión demuestran que el sistema utiliza efectivamente el rango completo de evaluación.

Resultados de Validación TSA

En junio de 2025, completamos una validación exhaustiva del TSA procesando **2,498 candidatos sintéticos** con una **tasa de éxito del 100%**.

Validación de Ausencia de Sesgos

Los resultados demuestran de manera concluyente la **ausencia de sesgos significativos** en todas las dimensiones analizadas:

Categoría	Tipo de Sesgo	Effect Size
Sesgos Demográficos		
Género	Negligible	0.018
Edad	Prácticamente nulo	0.002
Educación	Negligible	0.052
Etnia	Negligible	0.003
Sesgos Étnico-Culturales		
Origen Cultural	Negligible	0.003
Acento	Negligible	0.014
Origen del Nombre	Negligible	0.009
Sesgos Socioeconómicos		
Prestigio Universitario	Negligible	0.068
Tipo de Empresa	Negligible	0.044
Sesgos Cognitivos (Diferenciación Legítima)		
Estilo de Comunicación	Diferenciación intencional	0.744
Nivel de Confianza	Diferenciación intencional	0.707

Interpretación: Los valores de sesgos demográficos, étnico-culturales y socioeconómicos están muy por debajo del umbral establecido por Cohen ($d < 0.2$) para considerar un efecto como negligible. Los sesgos cognitivos muestran diferenciación intencional y legítima, ya que el estilo de comunicación y el nivel de confianza son competencias profesionales relevantes.

Consistencia Inter-variacional

El sistema demostró una **consistencia inter-variacional del 91.4%**, validando que las evaluaciones permanecen estables ante variaciones demográficas irrelevantes:

Posiciones Tech

92.6%

Posiciones Business

91.4%

Sector Tecnología

94.0%

Sector Salud

90.9%

Fiabilidad Test-Retest

La validación de consistencia temporal alcanzó una **puntuación test-retest del 97.9%**, demostrando una estabilidad temporal excepcional. Esta métrica confirma que el sistema produce evaluaciones altamente reproducibles cuando analiza perfiles idénticos en diferentes momentos.

Capacidad Discriminatoria

El análisis de 524 casos de "mismatch" validó la capacidad del sistema para diferenciar efectivamente entre candidatos aptos e inaptos. Las métricas de dispersión demuestran que el sistema utiliza efectivamente el rango completo de evaluación, con coeficientes de variación entre 7.4% y 11.1%, rangos de puntuación de 61-87 puntos (sobre 100), y detección exitosa tanto de candidatos claramente inadecuados (scores mínimos 10-34) como excepcionales (scores máximos 93-97).

Cumplimiento de Estándares Científicos

Estos resultados posicionan nuestro sistema significativamente por encima de los estándares mínimos requeridos para herramientas de evaluación en selección de personal:

- **Fiabilidad test-retest:** 97.9% (estándar mínimo: 80%)
- **Consistencia inter-evaluador:** 91.4% (estándar mínimo: 70%)
- **Ausencia de sesgos:** Todos los effect sizes < 0.1 (umbral aceptable: < 0.2)
- **Capacidad discriminatoria efectiva** con rangos amplios de evaluación

Esta validación empírica refuerza nuestra confianza en que **IRIA** proporciona evaluaciones justas, consistentes y reproducibles que apoyan decisiones de contratación equitativas basadas exclusivamente en méritos objetivos, **sin sesgo hacia la aprobación generalizada** de candidatos.

| 10. MANTENIMIENTO Y ACTUALIZACIÓN DEL SISTEMA DE IA

La IA no es una tecnología de "configurar y olvidar". Un sistema ético y eficaz requiere un mantenimiento y una supervisión constantes.

Programa de Actualización

Nuestros procesos de análisis y esquemas de evaluación se revisan y refinan continuamente. Este ciclo nos permite incorporar los últimos avances en la ingeniería de sistemas de IA, mejorar la precisión y responder a los cambios en el mercado laboral.

Supervisión de la Calidad

Mantenemos una supervisión activa sobre la calidad y consistencia de los análisis generados. Si bien no tenemos un

sistema de monitorización automatizado de los resultados en tiempo real, sí contamos con procesos internos para revisar periódicamente el rendimiento del sistema y la calidad de los informes.

Procedimiento ante Problemas

Si se detecta una anomalía o un cliente reporta un problema con un análisis, nuestro equipo técnico tiene el compromiso de investigarlo con la máxima prioridad. Nuestro objetivo es identificar la causa raíz, corregir cualquier problema en nuestro sistema y proporcionar una solución o un re-análisis al cliente afectado a la mayor brevedad posible.

| 11. PARTES INTERESADAS Y GESTIÓN DE RIESGOS

La implementación de IA en la selección de personal afecta a diversas partes interesadas. Gestionamos los riesgos asociados con la máxima diligencia.

Nuestros Clientes (Empresas Contratantes)

Proporcionamos una herramienta que aumenta su eficiencia y les ayuda a tomar decisiones más justas. El riesgo de una mala contratación se mitiga al posicionar a **IRIA** como una herramienta de apoyo que enriquece, pero no reemplaza, el juicio del reclutador.

Candidatos

Buscamos ofrecer una experiencia justa, accesible y respetuosa.

Gestión de Riesgos:

- **Temor a Sesgos Integrados:** Como se detalla en la sección de Pruebas, este es nuestro principal foco de mitigación. Al basar nuestro análisis 100% en texto y aplicar nuestro MEC, eliminamos clases enteras de sesgos relacionados con la apariencia física o el tono de voz.
- **Accesibilidad:** La plataforma está diseñada siguiendo las pautas de accesibilidad web. La entrevista se puede realizar completamente por voz o, si el candidato lo prefiere, interactuando únicamente a través del chat de texto.
- **Pérdida de Casos Excepcionales:** Reconocemos que un sistema automatizado puede no valorar adecuadamente a un candidato "atípico" pero brillante. Por eso, **IRIA** es una herramienta de apoyo y siempre animamos a revisar los perfiles completos y las grabaciones.

12. APÉNDICES

Anexo A: Detalles Técnicos de Validación TSA

Metodología de Consistencia Inter-variacional

La consistencia inter-variacional mide la estabilidad de las evaluaciones cuando se aplican variaciones demográficas irrelevantes a un mismo perfil base. Para cada grupo de variaciones, calculamos la desviación estándar de los scores y la convertimos a una métrica de consistencia utilizando la fórmula:

$$\text{Consistencia} = \max(0, 1 - (\text{desviación_estándar} / 15))$$

donde 15 representa el umbral máximo aceptable de variación en una escala de 100 puntos.

Resultados Detallados por Combinación Crítica (Test-Retest)

Las 30 combinaciones críticas evaluadas mostraron consistencias específicas que van desde 95.9% hasta 99.1%:

Posiciones Executive/Expert (≥98.0%):

- Tech Expert CTO: 98.0%
- Tech Expert Head: 98.7%
- Tech Senior Lead: 98.6%
- Business Senior VP: 98.4%
- Manufacturing Mid Supervisor: 99.1%
- Tech Junior QA: 99.1%

Posiciones Senior/Mid (97.0-98.0%):

- Business Expert CEO: 97.6%
- Healthcare Expert Cirujano/a: 97.2%
- Healthcare Expert Chief: 97.4%
- Tech Senior Arquitecto/a: 97.6%
- Business Senior Director/a: 97.3%

Distribución de Muestras por Fase TSA

- **Core Testing:** 1,035 análisis (41.4%) - Evaluaciones multi-sectoriales con variaciones demográficas
- **Bias Testing:** 709 análisis (28.4%) - Matriz completa de combinaciones demográficas
- **Mismatch Testing:** 524 análisis (21.0%) - Casos controlados de inadecuación candidato-posición
- **Consistency Testing:** 230 análisis (9.2%) - Repeticiones exactas para fiabilidad temporal

Criterios de Validación Utilizados

Estándares Internacionales de Referencia:

- American Psychological Association (APA) - Standards for Educational and Psychological Testing
- International Test Commission (ITC) Guidelines
- European Federation of Psychologists' Associations (EFPA) Standards

Umbral de Aceptación:

- Fiabilidad test-retest: ≥ 0.80 (ICC)
- Consistencia inter-evaluador: ≥ 0.70
- Effect size para sesgo: < 0.20 (Cohen's d)
- Poder estadístico: ≥ 0.80 ($1-\beta$)

Validación Estadística Completa

Todas las validaciones incluyen:

- Intervalo de confianza 95% para todas las métricas
 - Análisis de poder estadístico confirmado ($\beta > 0.80$)
 - Corrección por comparaciones múltiples aplicada
 - Validación cruzada estratificada por sector y nivel
-