

METODOLOGÍA DE EVALUACIÓN Y EVOLUCIÓN DE HABILIDADES

TOP QUANT ARENA

1. Objeto de la metodología

La competición Top Quant Arena no evaluará únicamente los resultados financieros obtenidos por cada participante. La competición tiene como objetivo adicional analizar la evolución de las capacidades necesarias para el diseño, evaluación y gestión de estrategias cuantitativas de inversión.

Para ello, cada participante dispondrá de un **Perfil de Habilidades Dinámico**, que evolucionará durante la competición a partir de evidencias objetivas obtenidas de su participación.

El sistema permitirá medir simultáneamente:

- El nivel estimado de cada participante en diferentes áreas de conocimiento y capacidad.
- La evolución de dichas capacidades durante la competición.
- Los conocimientos y habilidades del participante frente a otros participantes.
- La evolución de su perfil respecto al perfil de origen.
- La aproximación o alejamiento respecto de diferentes perfiles profesionales o *clusters*.
- La velocidad de aprendizaje y evolución.
- La relación entre las diferentes capacidades y el resultado obtenido en la competición.

La metodología no pretende determinar de forma absoluta el conocimiento o las capacidades de una persona. Las puntuaciones representan una **estimación dinámica de las capacidades demostradas durante la competición a partir de evidencias observables**.

2. Principios de la metodología

La metodología se fundamenta en los siguientes principios:

2.1. El perfil inicial no determina el resultado esperado

El perfil profesional o académico declarado por un participante únicamente establece su **punto de partida estadístico**.

La competición no presupone que un financiero deba obtener mejores resultados que un ingeniero, ni que un matemático deba superar a un programador.

El sistema no asignará un resultado financiero esperado en función del perfil profesional.

Un participante podrá evolucionar en cualquier dirección independientemente de su perfil inicial.

2.2. Las habilidades se estiman a partir de evidencias

Las puntuaciones de habilidades no se modificarán simplemente porque un participante obtenga una determinada rentabilidad.

El sistema utilizará diferentes tipos de evidencias:

1. **Resultados de inversión.**
2. **Comportamiento observable de la cartera.**
3. **Respuesta ante acontecimientos y cambios de mercado.**
4. **Micro-retos de conocimiento y razonamiento.**
5. **Rendimiento relativo frente a otros participantes.**
6. **Evolución del comportamiento entre diferentes fases de la competición.**

Cada evidencia podrá afectar a una o varias dimensiones de habilidades.

2.3. El sistema mide capacidad relativa y evolución

La metodología combinará dos conceptos diferentes:

Skill Rating: estimación del nivel demostrado en una determinada habilidad.

Competitive Rating: estimación de la capacidad competitiva del participante frente a otros participantes.

El primero describe **qué capacidades demuestra** el participante.

El segundo describe **cómo compite frente a otros participantes**.

2.4. El sistema se actualiza progresivamente

Las puntuaciones no se modificarán de forma radical a partir de una única observación.

La metodología utilizará un sistema de actualización progresiva inspirado en los sistemas de rating utilizados en competiciones como el ajedrez, adaptado a una competición multidimensional de inversión.

Una actuación extraordinaria frente a la expectativa estadística generará una actualización mayor que una actuación que se encuentre dentro de lo esperado.

2.5. La información se acumula

Cada nueva evidencia se incorpora a la información previamente disponible sobre el participante.

A medida que aumenta el número de observaciones, aumenta también la estabilidad de la estimación de sus habilidades.

Por este motivo, la velocidad de modificación de una habilidad será mayor al comienzo de la competición y se reducirá progresivamente a medida que el sistema disponga de más información.

3. Dimensiones de habilidades

El Perfil de Habilidades Dinámico estará compuesto por doce dimensiones:

Dimensión	Qué mide
Creatividad & Adaptación	Capacidad para generar soluciones originales y adaptarse a nuevos problemas, mercados y entornos.
Conocimiento Financiero	Conocimiento de instrumentos, valoración, carteras, derivados, fondos, renta fija, etc.
Conocimiento de Mercados	Comprensión del funcionamiento de los mercados, creación de precios, liquidez, volatilidad, microestructura y comportamiento de los participantes.
Matemáticas	Capacidad de razonamiento matemático, cálculo, álgebra, optimización y modelización cuantitativa.
Estadística	Probabilidad, inferencia, series temporales, estimación, contrastes y análisis de incertidumbre.
Data Analytics	Capacidad para obtener, limpiar, explorar, transformar e interpretar datos.
AI & Machine Learning	Capacidad para desarrollar y utilizar modelos de IA y machine learning.
Programación	Capacidad para convertir problemas y modelos en código y desarrollar algoritmos.
Diseño de Estrategias	Capacidad para transformar una hipótesis de inversión en una estrategia cuantitativa.
Quant research	Capacidad para investigar, validar y evaluar estrategias mediante backtesting, out-of-sample, robustez, etc.
Gestión del Riesgo	Capacidad para medir, controlar y gestionar riesgo, drawdown, exposición, sizing y capital.
Implementación & Producción	Capacidad para llevar una estrategia/modelo a producción, automatizarla, ejecutarla y monitorizarla.

Cada dimensión tendrá inicialmente una puntuación comprendida entre 0 y 10.

Las dimensiones se descomponen en diferentes competencias específicas, que podrán ser utilizadas para generar evidencias y micro-retos.

4. Composición de las dimensiones

Conocimiento financiero	Conocimiento de mercados	Matemáticas y razonamiento cuantitativo
Valoración de activos Renta fija Renta variable Fondos de inversión Derivados Gestión de carteras Optimización de portafolios	Formación de precios Liquidez Volatilidad Correlaciones Ciclos de mercado Comportamiento de los inversores Microestructura Ejecución	Cálculo Álgebra Optimización Probabilidad Procesos estocásticos Análisis numérico Razonamiento matemático

	Impacto de mercado	
Estadística y modelización Inferencia estadística Regresión Series temporales Análisis de distribuciones Estimación Hipótesis Modelización predictiva Validación estadística	Data Analytics Limpieza Transformación Exploración Detección de patrones Feature engineering Análisis de grandes datasets Tratamiento de datos incompletos Análisis de la calidad de los datos	Inteligencia Artificial Fundamentos de IA Aprendizaje supervisado & no supervisado Aprendizaje por refuerzo Modelos generativos y agenticos Optimización de modelos Ensemble learning Transfer learning Interpretabilidad de modelos Evaluación y validación de modelos
Programación Lógica y pensamiento computacional Algoritmos y estructuras de datos Lenguajes de programación Desarrollo de código Depuración y optimización Automatización y scripting	Diseño de algoritmos de inversión Diseño de estrategias Selección de variables Construcción de señales Asignación de capital Selección de activos Optimización Implementación de estrategias	Quant research Backtesting Validación fuera de muestra Control de overfitting Sesgo de supervivencia Look-ahead bias Robustez Análisis de sensibilidad Generación de datos sintéticos Generación de features sintéticas
Gestión del riesgo Volatilidad Drawdown VaR Stress testing Diversificación Correlaciones Sizing Exposición Gestión del capital Riesgo de modelo	Implementación y producción Deployment Automatización APIs Ejecución Monitorización Latencia Gestión de errores Escalabilidad MLOps DataOps Puesta en producción	Creatividad y adaptación Capacidad para formular hipótesis originales Encontrar nuevas fuentes de información Diseñar soluciones no convencionales Adaptarse a nuevos mercados Aprender rápidamente Cambiar de estrategia Combinar disciplinas diferentes

5. Perfil inicial

Cada participante comenzará la competición con un **Perfil Inicial de Habilidades**. Este perfil se determinará a partir del *cluster* de origen al que pertenezca el participante.

Cada *cluster* tendrá asociado un vector inicial de doce puntuaciones.

Estas puntuaciones **no constituyen una valoración definitiva del participante**. Representan exclusivamente el punto de partida estadístico utilizado por el sistema.

Perfil	Conocimient o Financiero	Conocimiento de Mercados	Matemáticas	Estadíst ica	Data Analytics	AI & ML	Programaci ón	Diseño Estrategias	Quant Research	Gestión del Riesgo	Implementación & Producción	Creatividad & Adaptación
Financieros	8	6	6	6	5	4	3	7	6	8	3	8
Traders	8	9	6	7	5	4	4	8	6	9	4	7
Ingenieros	5	5	9	7	6	6	8	4	7	4	8	8
Matemáticos	5	3	10	9	6	7	7	6	6	5	4	9
Físicos	5	3	9	9	7	7	7	6	6	5	4	9
Informáticos	4	3	7	6	7	8	10	7	6	5	9	8
Telecos	5	4	8	8	7	7	9	6	6	5	9	8
Comunidad Python	6	4	6	6	8	8	9	6	7	4	8	9
Profesores de IA	4	3	7	8	7	9	8	5	5	4	6	8
Profesores de Finanzas	9	7	6	7	5	3	3	7	6	9	3	7
Estudiantes de finanzas o IA	5	3	5	6	5	5	6	4	3	4	4	8
Data Engineers	4	3	7	6	9	7	9	3	4	4	10	8
Data Scientists	7	5	8	9	9	8	8	6	5	4	6	8
ML Engineers	5	3	8	8	9	9	9	6	5	4	9	8

6. Perfil individual dinámico

A partir del comienzo de la competición, cada participante dispondrá de un vector individual de habilidades:

$$S = (S_1, S_2, \dots, S_{12})$$

Cada componente representa la estimación actual de una de las doce dimensiones.

Por ejemplo:

AI & ML: 4,0

significa que el sistema estima inicialmente que el participante presenta un nivel 4,0 en esa dimensión.

Tras acumular evidencias, podría evolucionar a:

AI & ML: 5,2

y posteriormente:

AI & ML: 6,7.

La puntuación podrá aumentar o disminuir durante la competición.

7. Evidencias utilizadas para actualizar las habilidades

El sistema utilizará cinco grandes fuentes de evidencia.

7.1. Evidencias de comportamiento inversor

La plataforma analizará las decisiones de inversión observables, incluyendo, cuando resulte aplicable:

- Composición de la cartera.
- Concentración.
- Diversificación.
- Exposición.
- Frecuencia de rebalanceo.
- Turnover.
- Estabilidad de las posiciones.
- Variación de pesos.
- Reacción ante cambios de mercado.
- Gestión de pérdidas.
- Gestión del riesgo.
- Comportamiento ante cambios de régimen.
- Consistencia de las decisiones.

El sistema no necesita disponer del código utilizado por el participante para utilizar estas evidencias.

La metodología evaluará exclusivamente la información observable que genere la participación en la competición.

7.2. Evidencias de resultado

Se utilizarán las métricas oficiales definidas para cada fase.

Dependiendo de la fase, podrán incluir:

- Rentabilidad.
- Rentabilidad ajustada al riesgo.
- Volatilidad.
- Drawdown.
- Estabilidad.
- Consistencia.
- Concentración.
- Riesgo.
- Robustez.
- Adaptación.

Las métricas de resultado **no se interpretarán individualmente como una habilidad concreta**.

Por ejemplo, una elevada rentabilidad no implicará automáticamente una mayor puntuación en Matemáticas, Programación o Gestión del Riesgo.

Los resultados se utilizarán conjuntamente con otras evidencias.

8. Micro-retos

Durante la competición se podrán realizar **micro-retos** destinados a obtener evidencias directas sobre determinadas capacidades.

Los micro-retos podrán evaluar, entre otras:

- Conocimiento financiero.
- Conocimiento de mercados.
- Matemáticas.
- Estadística.
- Análisis de datos.
- Inteligencia Artificial.
- Programación.
- Diseño de estrategias.
- Quant research.
- Gestión del riesgo.

- Razonamiento.
- Creatividad y adaptación.

Los micro-retos podrán adoptar diferentes formatos:

- Preguntas de conocimiento.
- Resolución de problemas.
- Análisis de datos.
- Identificación de errores.
- Interpretación de resultados.
- Selección de modelos.
- Diseño de soluciones.
- Problemas cuantitativos.
- Ejercicios de programación.
- Análisis de estrategias.

Cada micro-reto tendrá un nivel de dificultad determinado por el sistema.

9. Rating de dificultad de los micro-retos

Los micro-retos utilizarán un sistema de dificultad equivalente a un *rating*.

Un reto podrá, por ejemplo, presentar una dificultad de:

Rating = 1.500

mientras que otro podrá presentar:

Rating = 2.000.

La dificultad será calibrada progresivamente a partir del comportamiento observado de los participantes.

De esta manera, el sistema podrá determinar no solamente si un participante responde correctamente, sino **qué nivel de dificultad es capaz de superar**.

Un participante que resuelva correctamente un reto considerablemente superior a su rating generará una evidencia de evolución mayor que otro participante para quien dicho reto resulte sencillo.

10. Competitive Rating

Además del Skill Rating de cada dimensión, el sistema calculará un **Competitive Rating**.

Este rating estará inspirado en los sistemas ELO utilizados en competiciones de resultado relativo, pero adaptado a una competición financiera en la que no existe necesariamente un enfrentamiento directo entre dos participantes.

El sistema no asumirá que un participante de un determinado perfil deba obtener una rentabilidad concreta.

En su lugar, estimará la probabilidad de que un participante supere a otros participantes o grupos de participantes en una determinada ventana temporal.

11. Competición contra distribuciones de participantes

Cada participante será comparado estadísticamente con diferentes grupos de referencia.

Entre otros:

11.1. Competición intra-cluster

Comparación con participantes pertenecientes a su mismo perfil de origen.

Ejemplo:

Financiero A frente al conjunto de participantes del cluster Financiero.

11.2. Competición cross-cluster

Comparación con otros perfiles.

Por ejemplo:

- Financiero A frente al conjunto de Ingenieros.
- Financiero A frente al conjunto de Matemáticos.

11.3. Competición global

Comparación con el conjunto de participantes de la competición.

11.4. Competición de élite

Comparación con determinados segmentos superiores de la población:

- Top 50%;
- Top 25%;
- Top 10%;
- Top 5%;
- Top 1%;

U otros umbrales que se establezcan para cada fase.

12. Percentiles competitivos

El sistema utilizará percentiles para expresar la posición relativa de un participante.

Por ejemplo:

Percentil 90: significa que el participante ha obtenido un resultado superior al de aproximadamente el 90% de la población de referencia.

Los percentiles podrán calcularse respecto de:

- Toda la población.
- Su cluster.
- Otros clusters.
- Participantes con características similares.
- Determinados grupos de referencia.

13. Probabilidad de superar un percentil

A partir del Competitive Rating y de la distribución histórica de resultados, el sistema podrá estimar probabilidades de superar diferentes umbrales.

Por ejemplo:

Objetivo	Probabilidad estimada
Superar P50 de Financieros	99%
Superar P75 de Financieros	91%
Superar P90 de Financieros	62%
Superar P95 de Financieros	37%
Superar P99 de Financieros	11%

Estas probabilidades constituyen estimaciones estadísticas, no garantías de resultados futuros.

14. Actualización tipo ELO

La actualización del Competitive Rating se realizará siguiendo una lógica equivalente a un sistema ELO.

La lógica general será:

$$\text{Rating nuevo} = \text{Rating anterior} + K \times (\text{Resultado observado} - \text{Resultado esperado})$$

donde:

- **Rating anterior** es el rating acumulado hasta ese momento.
- **Resultado esperado** representa la probabilidad estimada de obtener el resultado observado.
- **Resultado observado** representa el resultado efectivamente obtenido.
- **K** determina la velocidad de actualización.

Cuando un participante obtenga un resultado superior al esperado, su rating aumentará.

Cuando obtenga un resultado inferior al esperado, su rating disminuirá.

Cuanto mayor sea la diferencia entre el resultado esperado y el resultado observado, mayor será la actualización.

15. Ejemplo del funcionamiento del ELO

Supongamos que, antes de una semana determinada, el sistema estima que un participante tiene una probabilidad del 55% de situarse por encima del percentil 90 de su cluster.

Durante esa semana el participante efectivamente supera dicho percentil.

La actualización será positiva.

Si, por el contrario, el sistema estimaba una probabilidad del 95% y el participante no consigue superar el percentil 90, la actualización será negativa.

Por tanto:

El sistema no premia únicamente obtener buenos resultados. Premia especialmente obtener resultados superiores a los que indicaba la información disponible anteriormente.

16. ELO individual y ELO por dimensiones

El Competitive Rating podrá calcularse a dos niveles.

Competitive Rating general

Representa la capacidad competitiva global del participante.

Competitive Rating por dimensión

Cuando exista evidencia suficiente, podrá calcularse un rating específico asociado a determinadas dimensiones.

Esto permitirá distinguir, por ejemplo, entre:

- Un participante con elevada capacidad competitiva financiera.

- Un participante con elevada capacidad competitiva en AI, Data y Programación.
-

17. Actualización semanal

El Perfil de Habilidades se actualizará con periodicidad semanal.

Cada semana se generará una nueva observación del participante.

La actualización podrá incorporar:

- Comportamiento de cartera.
- Resultados de la semana.
- Eventos de mercado.
- Micro-retos realizados.
- Posición relativa.
- Evolución respecto a semanas anteriores.
- Resultados acumulados.

La actualización semanal no implica que todas las dimensiones tengan que modificarse.

Una determinada habilidad podrá:

- Aumentar
 - Permanecer estable
 - Disminuir
-

18. Fórmula general de evolución de una habilidad

Para cada habilidad se calculará una variación semanal:

$$\Delta S_{i,t} = L_{i,t} \times E_{i,t} \times C_{i,t} \times P_{i,t}$$

donde:

- $\Delta S_{i,t}$ = cambio semanal de la habilidad i ;
- $L_{i,t}$ = factor de aprendizaje o actualización;
- $E_{i,t}$ = evidencia obtenida;
- $C_{i,t}$ = factor de confianza;
- $P_{i,t}$ = factor de fase.

La puntuación será:

$$S_{i,t+1} = S_{i,t} + \Delta S_{i,t}$$

La fórmula se aplicará respetando los límites establecidos para cada fase.

19. Composición de la evidencia semanal

Como referencia metodológica, la evidencia semanal podrá estructurarse de la siguiente manera:

Fuente	Peso de referencia
Performance / resultados	35%
Comportamiento observable	25%
Micro-retos	20%
Posición relativa / percentil	10%
Evolución del perfil / cluster	10%

Estos pesos representan la estructura metodológica general del sistema.

La aplicación concreta de cada componente dependerá de la información disponible y de la naturaleza de cada fase.

Una dimensión solo recibirá evidencia procedente de aquellas fuentes que tengan una relación metodológicamente justificada con dicha dimensión.

20. No todas las métricas afectan a todas las habilidades

Una métrica podrá contribuir a una o varias dimensiones, pero no se realizará una correspondencia automática entre una métrica financiera y una habilidad.

Por ejemplo:

Una reducción del drawdown podrá generar evidencia favorable sobre Gestión del Riesgo y Adaptación.

Pero no implicará automáticamente una mejora en:

- Matemáticas
- Programación
- AI & ML

Del mismo modo:

Una rentabilidad extraordinaria podrá generar evidencia de Diseño de Estrategias, pero no demostrará por sí misma una determinada capacidad matemática o de programación.

La relación entre cada evidencia y cada dimensión está definida mediante una **Matriz de Evidencias**.

21. Matriz de Evidencias

La evolución de las habilidades de cada participante se determinará a partir de las evidencias observables generadas durante la competición.

El sistema no inferirá una habilidad a partir de una única métrica ni asumirá que un determinado resultado financiero implica necesariamente una determinada capacidad técnica.

Cada observación será procesada por una Matriz de Evidencias, que establece qué información aporta cada comportamiento observable sobre las distintas dimensiones de habilidad.

La matriz constituye el mecanismo que conecta el comportamiento observado de un participante con su perfil de habilidades.

21.1. Estructura de una evidencia

Cada tipo de evidencia tendrá asociados los siguientes elementos:

1. Una o varias dimensiones potencialmente afectadas.

Una misma evidencia podrá proporcionar información sobre más de una dimensión.

2. Un peso de contribución.

Determinará la importancia relativa de la evidencia para la estimación de cada dimensión.

3. Una dirección de efecto.

La evidencia podrá generar una señal positiva, negativa o neutra sobre una determinada habilidad.

4. Un nivel de intensidad.

Determinará la magnitud potencial de la señal generada por la evidencia.

5. Un nivel de confianza.

Representará el grado de fiabilidad con el que la evidencia permite realizar una inferencia sobre una habilidad.

6. Una dificultad o capacidad requerida.

Permitirá distinguir entre comportamientos que pueden producirse con conocimientos básicos y comportamientos que requieren capacidades más avanzadas.

7. Una persistencia temporal, cuando resulte necesaria.

Algunas evidencias deberán mantenerse durante varias observaciones antes de producir un efecto significativo sobre una habilidad.

8. **Una métrica de observación.**

Cada evidencia estará vinculada a una o varias métricas objetivamente calculables a partir de los datos disponibles.

9. **Una referencia de comparación.**

Cuando resulte apropiado, la evidencia será interpretada en términos relativos respecto al mercado, al benchmark, a otros participantes humanos o a la distribución histórica del propio participante.

21.2. Evidencia, señal y actualización

La metodología distingue expresamente entre tres conceptos:

Evidencia → Señal de habilidad → Actualización del Skill Rating

La **evidencia** es el hecho observable.

Por ejemplo: Un participante mantiene durante ocho semanas un drawdown situado entre los mejores de su población humana, sin asumir una concentración excesiva.

La **señal de habilidad** es la interpretación estadística de dicha evidencia.

En este caso podría existir una señal positiva relacionada con **Gestión del Riesgo**.

La **actualización del Skill Rating** es el cambio que finalmente experimenta la puntuación de dicha habilidad.

Por tanto, una evidencia no implica automáticamente una modificación fija de puntos.

El sistema no aplicará reglas del tipo:

"Si ocurre X, se suman 0,5 puntos."

La magnitud de la actualización dependerá de la combinación de evidencia, intensidad, confianza, posición relativa, persistencia temporal, historial previo y estado actual de la habilidad.

21.3. Dimensiones afectadas

Las evidencias podrán afectar a una o varias de las doce dimensiones de evaluación:

1. Conocimiento Financiero.
2. Conocimiento de Mercados.

3. Matemáticas y Razonamiento Cuantitativo.
4. Estadística y Modelización.
5. Data Analytics.
6. Inteligencia Artificial y Machine Learning.
7. Programación.
8. Diseño de Algoritmos de Inversión.
9. Quant Research.
10. Gestión del Riesgo.
11. Implementación y Producción.
12. Creatividad y Adaptación.

La matriz no presupone que una evidencia afecte necesariamente a todas las dimensiones relacionadas conceptualmente con ella.

La contribución estará determinada por la fuerza con la que dicha evidencia permita realizar una inferencia sobre cada habilidad.

21.4. Matriz de contribución

Cada evidencia se representará internamente mediante un vector de contribución sobre las doce dimensiones.

De forma simplificada:

Evidencia → [Finanzas, Mercados, Matemáticas, Estadística, Data, AI, Programación, Estrategia, Research, Riesgo, Producción, Creatividad]

Cada componente podrá tener:

- Contribución positiva
- Contribución negativa
- Contribución nula

La magnitud de cada contribución dependerá del peso establecido en la matriz y de las características concretas de la observación.

De esta manera, una misma evidencia podrá aportar información significativa sobre una habilidad, información moderada sobre otra y prácticamente ninguna información sobre una tercera.

21.5. Principales familias de evidencias

La matriz se organizará en nueve grandes familias:

Código	Familia de evidencia	Descripción
E1	Construcción de cartera	Selección de activos, pesos, concentración y diversificación
E2	Gestión dinámica	Rebalanceos, cambios de exposición y gestión temporal
E3	Performance	Resultados absolutos y ajustados por riesgo
E4	Comportamiento de mercado	Respuesta ante distintos regímenes y condiciones de mercado

E5	Gestión del riesgo	Volatilidad, drawdown, concentración, exposición y pérdidas extremas
E6	Micro-retos	Pruebas específicas de conocimientos y capacidades
E7	Adaptación entre fases	Capacidad de transferir conocimientos a nuevos entornos
E8	Rendimiento relativo / ELO	Capacidad competitiva relativa dentro de la población humana
E9	Persistencia temporal	Consistencia y estabilidad de los comportamientos observados

Estas familias no constituyen puntuaciones independientes.

Son categorías de evidencias utilizadas por el Skill Engine para construir las estimaciones de las distintas habilidades.

21.6. Evidencias procedentes de la cartera

Los participantes enviarán las posiciones o pesos de su cartera, de acuerdo con las reglas específicas de cada fase.

El sistema podrá observar, entre otros elementos:

- Selección de activos
- Concentración
- Diversificación
- Rotación
- Frecuencia de rebalanceo
- Cambios de exposición
- Estabilidad de las posiciones
- Comportamiento de la cartera
- Respuesta ante movimientos del mercado
- Relación entre riesgo asumido y resultado obtenido

Estas observaciones constituyen evidencia indirecta sobre determinadas capacidades de inversión.

Sin embargo, la mera obtención de una rentabilidad elevada no será interpretada automáticamente como evidencia de una determinada habilidad técnica.

El sistema tendrá en cuenta el contexto en el que se obtuvo el resultado, el riesgo asumido, la persistencia del comportamiento y su posición relativa respecto de otros participantes.

21.7. Normalización de las evidencias

Las métricas brutas no se utilizarán directamente para modificar las habilidades cuando su interpretación dependa de la escala o de las condiciones de mercado.

Cuando sea apropiado, las observaciones serán transformadas a métricas comparables mediante:

- Percentiles
- Deciles

- Rankings
- Normalización respecto a la distribución de participantes
- Comparación frente a benchmarks
- Comparación respecto al propio historial
- Medidas de persistencia

Por ejemplo, un drawdown del 10 % no tendrá necesariamente la misma interpretación en dos periodos de mercado diferentes.

El sistema podrá utilizar, por tanto, tanto el valor absoluto como la posición relativa de la observación.

Una evidencia se representa conceptualmente como:

Valor observado + posición relativa + contexto + persistencia

Esto permite que el sistema evalúe el comportamiento de forma contextualizada.

21.8. Comparación con la población humana

Las comparaciones de rendimiento utilizadas para la estimación ELO se realizarán exclusivamente entre **participantes humanos**.

Los agentes de Inteligencia Artificial y los inversores aleatorios participarán en la competición y sus resultados podrán utilizarse para otras métricas, comparaciones y análisis públicos de la competición, pero no formarán parte de la población utilizada para calcular el ELO de los participantes humanos.

Esta separación responde a una consideración metodológica:

- Los humanos
- Los agentes de IA
- Los inversores aleatorios

Constituyen poblaciones con mecanismos de generación de decisiones diferentes.

Por tanto, no resulta apropiado utilizarlos indistintamente como oponentes dentro de un mismo sistema de rating relativo de habilidades humanas.

21.9. ELO de rendimiento relativo

El sistema incorporará un mecanismo de rating relativo inspirado en el principio ELO utilizado en sistemas competitivos como el ajedrez.

El objetivo no será determinar si un participante ha obtenido un resultado "bueno" o "malo" en términos absolutos.

El objetivo será estimar su **capacidad competitiva relativa** frente a otros participantes humanos.

En lugar de establecer expectativas diferentes para cada perfil de participante, el sistema no asumirá que un financiero, ingeniero, matemático o programador deba obtener un determinado rendimiento.

Todos los participantes humanos competirán bajo las mismas condiciones.

El rating permitirá estimar, entre otras magnitudes:

- La probabilidad de que un participante supere a otro participante humano.
- La probabilidad de superar a participantes de un determinado nivel de rating.
- La probabilidad de situarse por encima de determinados percentiles de su población.
- La consistencia de esa capacidad competitiva a lo largo del tiempo.

El sistema podrá utilizar estas estimaciones como evidencia adicional para determinadas dimensiones, especialmente aquellas relacionadas con:

- Diseño de Algoritmos de Inversión.
- Quant Research.
- Gestión del Riesgo.
- Conocimiento de Mercados.
- Creatividad y Adaptación.

El ELO no constituirá por sí mismo una puntuación de habilidad.

Será una fuente de evidencia relativa utilizada por el Skill Engine.

21.10. ELO y clusters de participantes

La metodología podrá utilizar simultáneamente diferentes poblaciones de comparación.

En particular, podrán analizarse:

ELO intra-cluster

Compara el rendimiento de un participante con otros participantes pertenecientes a un mismo cluster de comportamiento o perfil.

Permite responder a preguntas como:

¿Qué probabilidad tiene este participante de superar a participantes similares a él?

ELO inter-cluster

Compara el rendimiento con participantes pertenecientes a otros clusters.

Permite analizar:

¿Qué capacidad tiene este participante para competir con participantes de perfiles diferentes?

El cluster no determinará una expectativa de rendimiento.

Su función será únicamente establecer una población de comparación estadísticamente útil.

Un participante podrá cambiar de cluster durante la competición si su comportamiento evoluciona de forma suficientemente significativa.

21.11. Intensidad de la evidencia

La intensidad representa la magnitud con la que una observación puede modificar una estimación de habilidad.

Se utilizarán cinco niveles:

Nivel	Interpretación
Muy baja	Evidencia débil o difícil de distinguir del comportamiento aleatorio
Baja	Evidencia limitada
Media	Evidencia relevante pero no concluyente
Alta	Evidencia fuerte
Muy alta	Evidencia excepcionalmente informativa

La intensidad no dependerá únicamente del resultado obtenido.

También podrá depender de:

- Magnitud de la desviación respecto de la población
- Dificultad del comportamiento
- Duración
- Condiciones de mercado
- Número de observaciones
- Consistencia
- Rareza estadística del comportamiento

21.12. Confianza de la evidencia

La confianza representa el grado en que una determinada observación permite atribuir una capacidad concreta.

Se distinguirá entre:

- Baja
- Media
- Alta
- Muy alta

Una evidencia puede tener una intensidad elevada y, simultáneamente, una confianza limitada.

Por ejemplo, un resultado extraordinariamente positivo durante un periodo muy corto puede constituir una observación de gran magnitud, pero no necesariamente permite concluir con elevada confianza que el participante posee una determinada habilidad.

Por el contrario, un comportamiento moderadamente positivo observado de manera consistente durante un periodo prolongado puede generar una evidencia de mayor confianza.

21.13. Dificultad de la evidencia

Cada evidencia podrá incorporar un nivel de dificultad o capacidad requerida.

Este parámetro permite distinguir entre comportamientos que pueden producirse con relativa facilidad y comportamientos que requieren una combinación de capacidades avanzadas.

La dificultad será especialmente relevante en la interpretación de los micro-retos y de determinadas decisiones de inversión.

No se utilizará para otorgar automáticamente una mayor puntuación.

Su función será determinar la capacidad informativa de una evidencia respecto del nivel de habilidad que pretende medir.

21.14. Persistencia temporal

El Skill Engine incorporará el principio de **persistencia de la evidencia**.

Una observación aislada no deberá producir, salvo circunstancias excepcionales, una modificación sustancial de una habilidad.

Cuando una evidencia sea susceptible de ruido, aleatoriedad o variación temporal, el sistema exigirá su repetición o persistencia antes de producir una actualización significativa.

La persistencia podrá medirse mediante:

- Número de semanas
- Número de observaciones
- Número de periodos de mercado
- Número de decisiones
- Consistencia de la señal

De esta manera se distinguirá entre:

- Un comportamiento excepcional
- Un patrón de comportamiento persistente

La persistencia podrá aumentar progresivamente la confianza de la evidencia.

21.15. Evitar la doble contabilización

Una misma observación podrá afectar a varias habilidades, pero no podrá contabilizarse como si fueran evidencias independientes.

El sistema aplicará un mecanismo de **normalización de contribuciones cruzadas**.

Por ejemplo, una decisión de reducir exposición antes de un episodio de elevada volatilidad podría generar información sobre:

- Conocimiento de Mercados
- Diseño de Estrategias
- Gestión del Riesgo
- Creatividad y Adaptación

Sin embargo, las cuatro contribuciones procederán de una misma observación subyacente.

Por ello, la matriz distribuirá la capacidad informativa de la evidencia entre las distintas dimensiones en lugar de multiplicarla artificialmente.

Este mecanismo evita que una única decisión pueda provocar una actualización desproporcionada del perfil completo del participante.

21.16. Independencia de las evidencias

Cuando varias métricas procedan del mismo fenómeno subyacente, el sistema evitará tratarlas como evidencias completamente independientes.

Por ejemplo:

- Rentabilidad
- Sharpe
- Volatilidad
- Drawdown

Están parcialmente relacionadas.

La metodología tiene en cuenta estas dependencias para evitar que un único fenómeno sea contabilizado repetidamente.

La cantidad de información aportada por un conjunto de métricas estará limitada por la información realmente independiente que contienen.

21.17. Matriz de contribución

A modo de referencia, las principales relaciones entre familias de evidencia y dimensiones serán las siguientes:

Familia	Financiero	Mercados	Matemáticas	Estadística	Data	AI/ML	Prog.	Estrategia	Research	Riesgo	Producción	Adaptación
E1 Cartera	Alta	Alta	Baja	Baja	Baja	Nula	Nula	Muy alta	Media	Alta	Baja	Media
E2 Gestión dinámica	Alta	Muy alta	Baja	Media	Baja	Nula	Nula	Muy alta	Alta	Muy alta	Media	Alta
E3 Performance	Media	Media	Nula	Media	Nula	Nula	Nula	Alta	Alta	Alta	Baja	Media
E4 Mercado	Media	Muy alta	Media	Alta	Media	Nula	Nula	Alta	Alta	Alta	Nula	Muy alta
E5 Riesgo	Alta	Alta	Media	Alta	Baja	Nula	Nula	Alta	Alta	Muy alta	Media	Alta
E6 Micro-retos	Según prueba	Según prueba	Según prueba	Según prueba	Según prueba	Según prueba	Según prueba	Según prueba	Según prueba	Según prueba	Según prueba	Según prueba
E7 Adaptación	Media	Alta	Media	Alta	Alta	Alta	Media	Alta	Alta	Alta	Alta	Muy alta
E8 ELO humano	Media	Alta	Baja	Media	Nula	Nula	Nula	Muy alta	Muy alta	Alta	Media	Alta
E9 Persistencia	Media	Alta	Media	Alta	Alta	Alta	Media	Alta	Muy alta	Alta	Alta	Muy alta

Los niveles anteriores representan la capacidad potencial de contribución de cada familia de evidencia, no una asignación directa de puntos.

La contribución efectiva está modulada por la intensidad, confianza, dificultad, persistencia y posición relativa de cada observación.

21.18. Micro-retos como evidencia directa

Los micro-retos constituirán una fuente especialmente importante de evidencia directa.

A diferencia de las evidencias derivadas del comportamiento de la cartera, los micro-retos permitirán evaluar determinadas capacidades mediante tareas diseñadas específicamente para ello.

Podrán utilizarse, entre otros, para observar:

- Razonamiento matemático
- Estadística
- Análisis de datos
- Inteligencia Artificial
- Programación
- Diseño algorítmico
- Investigación cuantitativa
- Gestión del riesgo
- Interpretación de mercados
- Resolución de problemas nuevos

Los micro-retos no tendrán necesariamente la misma ponderación en todas las fases.

Su dificultad podrá aumentar progresivamente, permitiendo observar niveles superiores de capacidad a medida que avance la competición.

21.19. Evidencias negativas

La matriz también podrá incorporar evidencias con dirección negativa.

Una evidencia negativa no significa necesariamente que un participante "no tenga" una determinada habilidad.

Significa que su comportamiento observado constituye información contraria a la hipótesis de que posee un nivel elevado de dicha capacidad.

Por ejemplo:

- Asumir un riesgo excesivo de forma persistente.
- Incrementar significativamente el turnover sin mejorar el resultado.
- Mantener una estrategia que deja de funcionar pese a disponer de información suficiente para identificar el cambio.
- Presentar una elevada inestabilidad en decisiones sucesivas.
- Mostrar errores persistentes en micro-retos.

Podrán generar señales negativas sobre determinadas dimensiones.

Las señales negativas estarán sujetas a los mismos mecanismos de confianza, persistencia y normalización que las positivas.

21.20. Ejemplos de evidencias

A modo ilustrativo, la matriz podrá contener evidencias como las siguientes:

Evidencia	Mercados	Estrategia	Research	Riesgo	Adaptación	Intensidad	Confianza
Reacción adecuada ante cambio de régimen	Alta	Alta	Media	Media	Muy alta	Alta	Alta
Control persistente del drawdown	Media	Media	Baja	Muy alta	Media	Alta	Muy alta
Exceso de turnover sin mejora de resultado	Media	Negativa	Media	Negativa	Negativa	Media	Alta
Robustez entre periodos	Media	Alta	Muy alta	Media	Alta	Alta	Alta
Adaptación exitosa a nuevo instrumento	Alta	Alta	Alta	Alta	Muy alta	Muy alta	Alta
Mejora persistente frente al propio historial	Media	Alta	Alta	Alta	Muy alta	Alta	Alta
Superación consistente de participantes humanos de mayor rating	Alta	Muy alta	Muy alta	Alta	Alta	Alta	Alta

Estos ejemplos son ilustrativos de la lógica de la matriz. La matriz completa de evidencias y sus parámetros está definida y será aplicada de forma homogénea a todos los participantes.

21.21. Aplicación por el Skill Engine

La Matriz de Evidencias no constituye por sí misma el sistema de puntuación.

Su función es proporcionar al **Skill Engine** una representación estructurada de la información obtenida durante la competición.

El proceso general será:

1. Observación: El sistema registra un comportamiento o resultado observable.

↓

2. Extracción de métricas: Se calculan las métricas correspondientes.

↓

3. Normalización: Las métricas se transforman cuando sea necesario mediante percentiles, rankings, benchmarks u otras medidas relativas.

↓

4. Identificación de evidencia: La observación se clasifica dentro de una o varias categorías de evidencia.

↓

5. Evaluación de la evidencia: Se determinan:

- Dirección
- Intensidad
- Confianza
- Dificultad
- Persistencia
- Dimensiones afectadas

↓

6. Actualización del Skill Rating: El Skill Engine incorpora la nueva información al historial del participante.

↓

7. Aplicación del Skill Ceiling: La puntuación publicada se limita al Skill Ceiling correspondiente a la fase.

↓

8. Actualización del perfil: Se genera el nuevo vector de habilidades del participante.

21.22. Principio fundamental

El principio fundamental de la Matriz de Evidencias es que: Una habilidad no se determina por un resultado aislado, sino por la acumulación de evidencias observables y estadísticamente interpretables a lo largo del tiempo.

El sistema buscará, por tanto, identificar patrones de comportamiento y no únicamente resultados.

Un participante podrá obtener una rentabilidad excepcional y no experimentar una mejora equivalente en todas sus habilidades.

Del mismo modo, podrá experimentar una mejora significativa en una determinada habilidad aunque su resultado financiero en un periodo concreto no sea excepcional, si existen evidencias adicionales que sustenten dicha evolución.

El objetivo del sistema es estimar progresivamente **qué capacidades posee el participante, con qué nivel de confianza y cómo están evolucionando**, y no simplemente asignarle una puntuación en función de su rentabilidad.

La Matriz de Evidencias constituye, de este modo, el puente metodológico entre el comportamiento observable de cada participante y la evolución de su perfil de habilidades.

22. Comparación con el comportamiento esperado del propio participante

El sistema tendrá en cuenta la trayectoria individual del participante.

No se utilizará únicamente la comparación con la población.

Por ejemplo, si un participante ha demostrado históricamente un comportamiento estable y posteriormente comienza a mostrar una deterioración significativa, dicha evolución podrá generar evidencia negativa.

De igual manera, una mejora persistente respecto al comportamiento histórico podrá generar evidencia positiva.

Esto permite distinguir entre:

- Una fluctuación puntual
- Una transformación persistente de comportamiento

23. Inercia del Skill Rating

Las puntuaciones tendrán una determinada inercia.

Una única actuación no podrá transformar radicalmente una habilidad.

La velocidad de actualización será mayor cuando:

- Exista poca información histórica

- La evidencia sea muy fuerte
- El resultado sea significativamente superior o inferior al esperado
- La dificultad de la evidencia sea elevada

La velocidad de actualización será menor cuando:

- Exista un historial amplio
- La evidencia sea débil
- El resultado se encuentre dentro de lo esperado
- La habilidad ya presente una elevada confianza estadística

24. Confianza de la estimación

Cada Skill Rating estará acompañado internamente de un nivel de confianza.

Por ejemplo:

AI & ML: 7,2 — Confianza 86%

La confianza dependerá, entre otros factores, de:

- Número de evidencias
- Diversidad de las evidencias
- Dificultad de las pruebas
- Consistencia temporal
- Estabilidad del comportamiento
- Cantidad de observaciones independientes

Una puntuación obtenida con pocas evidencias tendrá una confianza menor que una puntuación equivalente obtenida después de numerosos meses y diferentes tipos de evidencias.

25. Diferencia entre Skill Level y Skill Growth

La competición distinguirá entre:

Skill Level: Nivel actual estimado.

Skill Growth: Evolución respecto al nivel inicial.

Por ejemplo:

AI & ML

Inicial: 4,0

Actual: 7,1

Growth: +3,1

Esto permite diferenciar entre participantes con un elevado nivel inicial y participantes que hayan experimentado una evolución extraordinaria durante la competición.

26. Learning Velocity

El sistema calculará adicionalmente **Learning Velocity**, entendida como la velocidad a la que evoluciona el perfil de habilidades.

Por ejemplo:

Skill Growth = +2,4 puntos

Tiempo = 12 semanas

Learning Velocity = +0,20 puntos/semana

Esta métrica permitirá identificar participantes que, independientemente de su nivel inicial, presentan una elevada capacidad de aprendizaje y adaptación.

27. Profile Migration

El sistema también analizará la distancia entre el perfil individual de cada participante y los diferentes perfiles profesionales de referencia.

Cada cluster estará representado mediante un vector de doce habilidades.

Por ejemplo:

Financial Cluster

[8, 6, 6, 6, 5, 4, 3, 7, 6, 8, 3, 8]

Engineering Cluster

[5, 5, 9, 7, 6, 6, 8, 4, 7, 4, 8, 8]

ML Engineering Cluster

[5, 3, 8, 8, 9, 9, 9, 6, 5, 4, 9, 8]

El sistema calculará la distancia entre el perfil individual y dichos centroides.

28. Evolución hacia otros clusters

Un participante podrá aproximarse progresivamente a otro cluster.

Por ejemplo:

Semana 1

Financial: 82% de similitud

Engineering: 11%

Mathematical: 7%

Semana 20

Financial: 57%

Engineering: 29%

Mathematical: 14%

Semana 40

Financial: 39%

Engineering: 38%

Mathematical: 23%

Esta evolución se identificará como: **Profile Migration**

La migración no significa que el participante cambie oficialmente de profesión.

Significa que su conjunto de capacidades demostradas se está aproximando estadísticamente al patrón de capacidades de otro perfil.

29. Límites de las habilidades por fase

La competición se desarrolla a través de cuatro fases progresivas, cada una de ellas caracterizada por un mercado, unos instrumentos financieros, unas fuentes de información y unos problemas de inversión diferentes.

El sistema de evaluación no pretende asumir que todas las capacidades de un participante pueden ser observadas con la misma profundidad desde el inicio de la competición. Una determinada habilidad puede manifestarse parcialmente en una fase y disponer de mucha más evidencia observable en fases posteriores.

Por este motivo, cada una de las doce dimensiones de evaluación dispondrá, en cada fase, de un Skill Ceiling, o límite máximo de puntuación observable.

El Skill Ceiling representa el nivel máximo que el sistema puede atribuir justificadamente a una determinada habilidad en una fase concreta, teniendo en cuenta la naturaleza de las pruebas, los datos disponibles, las decisiones observables y las evidencias generadas hasta ese momento.

29.1. Naturaleza del Skill Ceiling

El Skill Ceiling no representa el nivel máximo de capacidad que puede tener un participante.

Representa exclusivamente el nivel máximo que el sistema considera metodológicamente justificable atribuirle en ese momento.

Por tanto, que una habilidad tenga un Skill Ceiling de 6,5 en una determinada fase no significa que el participante no pueda poseer una capacidad real superior. Significa que la competición todavía no ha generado suficiente evidencia para distinguir de forma suficientemente fiable entre distintos niveles superiores de dicha capacidad.

Esta distinción resulta especialmente relevante en aquellas dimensiones cuya observación directa es limitada durante las primeras fases.

Por ejemplo, los resultados obtenidos en una cartera financiera pueden proporcionar información relevante sobre el Diseño de Estrategias, la Gestión del Riesgo o el Conocimiento de Mercados, pero no permiten determinar por sí solos, con el mismo grado de confianza, la capacidad de un participante en Programación o Inteligencia Artificial.

Estas últimas dimensiones requerirán evidencias adicionales, obtenidas mediante micro-retos, pruebas específicas, comportamiento longitudinal y adaptación a problemas progresivamente más complejos.

29.2. Aplicación del Skill Ceiling

La puntuación calculada por el sistema para cada habilidad será denominada **Skill Rating**.

El Skill Rating se calculará a partir de las evidencias disponibles y de los mecanismos de actualización establecidos en la metodología de evolución de habilidades.

Una vez calculado el Skill Rating, se aplicará el Skill Ceiling correspondiente a la fase:

Skill Rating publicado = min(Skill Rating calculado, Skill Ceiling de la fase)

De esta forma, una puntuación nunca podrá superar el nivel máximo de observabilidad establecido para dicha fase.

El sistema conservará, no obstante, la evidencia generada y el nivel de capacidad estimado internamente aunque este se encuentre por encima del Skill Ceiling publicado.

Esto permite evitar que una capacidad demostrada prematuramente quede perdida o tenga que ser reconstruida desde cero en fases posteriores.

Por ejemplo, si un participante obtiene un Skill Rating estimado de 7,4 en una determinada habilidad durante una fase cuyo Skill Ceiling es 6,5, su puntuación publicada será 6,5. Sin embargo, la evidencia que sustenta el 7,4 permanecerá incorporada al historial del participante y podrá ser utilizada en las actualizaciones posteriores cuando el Skill Ceiling aumente.

El Skill Ceiling constituye, por tanto, un límite de observabilidad y no un límite de capacidad.

29.3. Progresión de los Skill Ceilings

Los Skill Ceilings aumentarán progresivamente a lo largo de las cuatro fases.

La progresión responde a tres principios:

1. **Profundidad de observación:** cuanto mayor sea la cantidad y variedad de evidencias disponibles sobre una habilidad, mayor será el nivel máximo que podrá atribuirse.
2. **Complejidad del entorno:** las fases posteriores introducen instrumentos, mercados y problemas de inversión más complejos, permitiendo observar capacidades que no podían evaluarse suficientemente en las primeras fases.
3. **Acumulación de evidencia:** las evidencias obtenidas en fases anteriores no desaparecen. El sistema construye un historial longitudinal de cada participante que permite evaluar no solo su nivel actual, sino también su evolución.

Los límites establecidos para las cuatro fases serán los siguientes:

Dimensión	Fase 1	Fase 2	Fase 3	Fase 4
Conocimiento Financiero	8,0	8,5	9,0	10,0
Conocimiento de Mercados	8,0	8,5	9,0	10,0
Matemáticas y Razonamiento Cuantitativo	6,5	7,0	8,0	10,0
Estadística y Modelización	7,0	7,5	8,5	10,0
Data Analytics	7,5	8,0	8,5	10,0
Inteligencia Artificial y Machine Learning	5,0	6,0	8,5	10,0
Programación	5,0	7,0	9,0	10,0
Diseño de Algoritmos de Inversión	7,5	8,5	9,0	10,0
Quant Research	7,5	8,0	9,0	10,0
Gestión del Riesgo	8,0	8,5	9,5	10,0
Implementación y Producción	7,0	8,0	9,0	10,0
Creatividad y Adaptación	6,0	7,0	8,5	10,0

La escala utilizada para todas las dimensiones será de 0 a 10.

El valor 10 representa el máximo nivel de habilidad que el sistema puede atribuir dentro del modelo de evaluación. No implica necesariamente un dominio absoluto o universal de la materia, sino el máximo nivel definido por la escala de la competición.

30. Límites específicos por fase

Los Skill Ceilings anteriores responden a la capacidad de cada fase para generar evidencia sobre las distintas dimensiones de conocimiento y habilidad.

30.1. Fase 1 · Inversión en renta variable

La primera fase se desarrolla sobre un índice de renta variable que será desvelado el primer día en el que se tenga acceso al backtesting (15 de diciembre), constituyendo la primera observación competitiva de todos los participantes.

En esta fase se dispone de información especialmente relevante sobre:

- Construcción de carteras
- Selección de activos
- Asignación de pesos
- Diversificación
- Concentración
- Exposición
- Volatilidad
- Drawdown
- Comportamiento frente a movimientos del mercado
- Rebalances
- Estabilidad de las decisiones
- Selección de estrategias
- Investigación cuantitativa aplicada
- Tratamiento y utilización de información disponible

Por este motivo, las dimensiones relacionadas directamente con la construcción y gestión de carteras podrán alcanzar niveles de observación relativamente elevados.

En cambio, otras capacidades, como Programación, Inteligencia Artificial o Matemáticas, tendrán un Skill Ceiling inferior, dado que el sistema no dispone de acceso al código ni puede inferir directamente dichas capacidades a partir del resultado financiero.

La Fase 1 constituye, por tanto, una observación especialmente significativa de la capacidad de un participante para construir y gestionar una cartera bajo condiciones reales de mercado, pero no pretende realizar todavía una evaluación completa de todas sus capacidades técnicas.

30.2. Fase 2 · Inversión en fondos de inversión

La segunda fase introduce un cambio sustancial respecto de la primera.

Los participantes deberán trabajar con fondos de inversión, donde la información disponible sobre la composición histórica será limitada (incompleta) y donde los cambios en la composición de la cartera implicarán periodos de transición y exposición temporal a diferentes condiciones de mercado.

Esta fase permite generar nuevas evidencias relacionadas con:

- Análisis de instrumentos financieros de otra naturaleza
- Gestión de información incompleta
- Gestión y tratamiento de grandes volúmenes de información
- Reconstrucción e interpretación de información histórica

- Comportamiento de los fondos
- Análisis de track records
- Gestión temporal
- Gestión de rebalanceos
- Coste de oportunidad
- Adaptación de estrategias
- Gestión del riesgo ante información imperfecta

Por ello, se incrementan los Skill Ceilings de las dimensiones relacionadas con análisis de información, adaptación, implementación y conocimiento financiero.

La Fase 2 permite, además, observar con mayor profundidad la capacidad del participante para tomar decisiones cuando la información disponible no es completa y cuando las decisiones de inversión tienen consecuencias temporales que no aparecen en una cartera de renta variable convencional.

30.3. Fase 3 · Inversión en algoritmos de inversión

La tercera fase modifica nuevamente la naturaleza del problema.

Los participantes ya no deberán seleccionar directamente acciones o fondos, sino que deberán seleccionar, asignar capital y gestionar inversiones sobre algoritmos desarrollados por otros traders. Invertirán en personas.

El participante dispondrá de información sobre el comportamiento histórico, track record y riesgo de cada algoritmo, pero no conocerá la lógica interna mediante la cual se generan sus decisiones.

Esta fase proporciona evidencia adicional especialmente relevante sobre:

- Evaluación de estrategias
- Análisis de track records
- Análisis de riesgo
- Riesgo de modelo
- Selección de traders
- Asignación de capital
- Diversificación entre estrategias
- Evaluación de persistencia
- Identificación de robustez
- Análisis de comportamiento
- Selección entre estrategias con diferentes perfiles de riesgo

En consecuencia, se incrementan de forma significativa los límites de Estadística y Modelización, Quant Research, Gestión del Riesgo, Diseño de Estrategias y Creatividad y Adaptación.

La Fase 3 constituye, por tanto, una prueba especialmente relevante de la capacidad de un participante para identificar inteligencia inversora en terceros, evaluar la robustez de una estrategia y asignar capital bajo incertidumbre.

30.4. Fase 4 · Derivados climáticos

La cuarta fase representa el mayor nivel de complejidad de la competición.

Los participantes deberán enfrentarse a un entorno financiero diferente de los mercados tradicionales, utilizando estrategias sobre derivados climáticos y otros instrumentos vinculados a variables ambientales y eventos catastróficos.

Se incorporarán datos procedentes de distintos países y variables relacionadas, entre otras, con:

- Temperatura
- Agua
- Viento
- Fenómenos climáticos
- Variables ambientales
- Eventos catastróficos
- Incertidumbre
- Diferentes horizontes temporales
- Derivados con vencimientos de 6, 9 y 12 meses

Esta fase permite observar por primera vez, de forma conjunta y con una profundidad significativamente superior, la capacidad del participante para **transferir conocimientos y habilidades a un dominio financiero nuevo**.

Se generarán evidencias relacionadas con:

- Derivados financieros
- Valoración y comportamiento de instrumentos no tradicionales
- Análisis estadístico de variables ambientales
- Modelización de incertidumbre
- Análisis de grandes datasets
- Identificación de relaciones entre variables
- Construcción de estrategias
- Gestión del riesgo
- Adaptación a nuevos mercados
- Transferencia de conocimiento
- Utilización de herramientas cuantitativas y de inteligencia artificial
- Creatividad
- Generación y evaluación de hipótesis
- Capacidad para resolver problemas que no se corresponden con los mercados tradicionales

Por primera vez, todas las dimensiones podrán alcanzar el nivel máximo de la escala.

El establecimiento de un Skill Ceiling de 10 en todas las dimensiones en la Fase 4 significa que, al alcanzar esta etapa, la competición dispone potencialmente de mecanismos suficientes para observar cualquier dimensión hasta el máximo de la escala.

Esto no implica que todos los participantes en esta fase alcancen una puntuación de 10.

Una puntuación de 10 únicamente podrá alcanzarse cuando las evidencias acumuladas y las reglas de actualización del Skill Engine justifiquen dicho nivel.

30.5. Interpretación de la evolución

Los Skill Ceilings no implican que las habilidades deban aumentar automáticamente de una fase a otra.

Un participante puede:

- Mejorar una habilidad
- Mantenerla estable
- Experimentar una reducción
- Mejorar significativamente unas dimensiones y empeorar otras
- Desarrollar capacidades inicialmente poco relacionadas con su perfil de partida
- Aproximarse progresivamente a otro perfil de referencia

La evolución de las habilidades dependerá exclusivamente de las evidencias generadas por su comportamiento, sus resultados relativos, los micro-retos y las demás fuentes de información contempladas en la metodología.

Por tanto, el incremento del Skill Ceiling entre fases no supone una bonificación automática.

Simplemente permite que el sistema reconozca niveles de habilidad que anteriormente no podía observar o atribuir con suficiente confianza.

La evolución individual será, en consecuencia, el resultado de la combinación de:

perfil inicial + evidencias observadas + rendimiento relativo + adaptación + persistencia temporal + evolución del Skill Ceiling.

De esta manera, la competición no medirá únicamente quién obtiene mejores resultados en un determinado momento, sino también **cómo evoluciona el conjunto de capacidades de cada participante a medida que aumenta la complejidad del entorno competitivo.**

31. La fase como prueba de transferencia

La transición entre fases tendrá además una función metodológica específica.

El sistema analizará la capacidad del participante para trasladar conocimientos y comportamientos adquiridos en un entorno a otro entorno diferente.

Esta capacidad será especialmente relevante para la dimensión de **Creatividad & Adaptación.**

Aunque también podrá generar evidencia adicional en:

- Estrategia
- Research

- Mercados
- Estadística
- Data
- AI
- Gestión del riesgo

Un participante que mantenga un elevado desempeño cuando cambia radicalmente el entorno de inversión demostrará una mayor capacidad de transferencia y adaptación que un participante cuyo rendimiento dependa de un único entorno.

32. Performance Outperformance

El sistema calculará una métrica denominada **Profile Outperformance**.

Esta métrica comparará:

- El rendimiento competitivo observado por un participante
- El rendimiento competitivo asociado estadísticamente a participantes con perfiles iniciales similares

Por ejemplo:

Rendimiento esperado respecto a su población de referencia: P68

Rendimiento observado: P91

Profile Outperformance: +23 percentiles

Esta métrica permitirá identificar participantes que superan significativamente las expectativas asociadas a su perfil de origen.

No se interpretará como una predicción ni como una medida absoluta de talento.

33. Ejemplo completo de evolución

Un participante comienza como Financiero.

Perfil inicial

Dimensión	Inicial
Financiero	8,0
Mercados	6,0
Matemáticas	6,0
Estadística	6,0
Data	5,0
AI & ML	4,0
Programación	3,0
Estrategia	7,0

Dimensión	Inicial
Research	6,0
Riesgo	8,0
Producción	3,0
Creatividad	8,0

Después de varios meses, las evidencias acumuladas pueden producir:

Dimensión	Actual
Financiero	8,3
Mercados	7,2
Matemáticas	6,5
Estadística	7,1
Data	6,8
AI & ML	6,3
Programación	5,4
Estrategia	8,0
Research	7,6
Riesgo	8,5
Producción	5,1
Creatividad	8,7

El participante continúa perteneciendo a su *cluster* de origen, pero su perfil de habilidades observado ha evolucionado.

La distancia respecto al cluster Financiero puede haber aumentado mientras que simultáneamente disminuye la distancia respecto a clusters como Engineering Quant o AI Quant.

34. Qué información se mostrará al participante

La plataforma mostrará, entre otras métricas:

- **Perfil de habilidades:** Las doce dimensiones y su puntuación actual.
- **Evolución:** Variación respecto al inicio de la competición.
- **Skill Growth:** Incremento o disminución acumulada.
- **Learning Velocity:** Velocidad de evolución.
- **Competitive Rating:** Capacidad competitiva estimada.
- **Percentiles:** Posición relativa frente a diferentes grupos.
- **Profile Migration:** Evolución respecto a los diferentes *clusters*.
- **Confidence:** Nivel de confianza asociado a las estimaciones.

La presentación podrá utilizar representaciones gráficas, rankings, evolución temporal y comparaciones entre perfiles.

35. Ranking de habilidades

Además del **Top Quant Arena**, la competición generará rankings específicos de habilidades.

Entre otros:

- Top Quant Financial Skill Ranking
- Top Quant Market Skill Ranking
- Top Quant Mathematics Ranking
- Top Quant Statistics Ranking
- Top Quant Data Ranking
- Top Quant AI Ranking
- Top Quant Programming Ranking
- Top Quant Strategy Ranking
- Top Quant Research Ranking
- Top Quant Risk Ranking
- Top Quant Production Ranking
- Top Quant Creativity & Adaptation Ranking

Estos rankings complementarán (no sustituirán) al ranking oficial de la competición.

36. Ranking de evolución

La plataforma publicará adicionalmente rankings relacionados con la evolución:

- **Top Quant Most Improved:** Participantes con mayor crecimiento absoluto de habilidades.
 - **Top Quant Fastest Learners:** Participantes con mayor velocidad de aprendizaje.
 - **Top Quant Profile Transformation:** Participantes que hayan experimentado una mayor transformación respecto a su perfil inicial.
 - **Top Quant Cross-Cluster:** Participantes que hayan demostrado capacidades significativamente superiores a las esperadas para su *cluster* de origen.
-

37. Principio de separación entre capacidad y resultado

La metodología distingue expresamente entre:

- Resultado financiero
- Capacidad demostrada

Un resultado financiero positivo o negativo puede estar condicionado por factores como:

- Condiciones de mercado
- Régimen de mercado
- Azar
- Concentración

- Exposición
- Acontecimientos extraordinarios
- Características específicas del instrumento

Por este motivo, ningún resultado aislado determinará por sí solo la evolución de una habilidad.

La modificación de las puntuaciones dependerá de la combinación de diferentes evidencias y de su persistencia temporal.

38. Tratamiento de resultados extraordinarios

Cuando un participante obtenga un resultado excepcionalmente elevado o reducido, el sistema podrá aplicar mecanismos de estabilización estadística.

Entre otros:

- Comparación con periodos anteriores
- Comparación con la distribución de participantes
- Comparación con participantes similares
- Análisis de persistencia
- Reducción de la influencia de observaciones aisladas
- Aumento progresivo de la confianza cuando el comportamiento se repita

El objetivo es evitar que una única observación extraordinaria produzca una transformación artificial del Perfil de Habilidades.

39. Tratamiento de información insuficiente

Cuando no exista evidencia suficiente para evaluar una determinada dimensión, el sistema podrá mantener su puntuación sin modificación o mostrarla con un nivel de confianza reducido.

La ausencia de evidencia no se interpretará como evidencia de ausencia de capacidad.

Por ejemplo, si durante una fase determinada un participante no tiene oportunidades suficientes para demostrar conocimientos avanzados de programación, su puntuación de Programación no se reducirá automáticamente por ese motivo.

40. Evolución positiva y negativa

Las habilidades podrán evolucionar en ambas direcciones.

Un participante podrá experimentar:

AI & ML: $4,0 \rightarrow 5,2 \rightarrow 6,8$

mientras simultáneamente:

Gestión del Riesgo: $8,0 \rightarrow 7,7 \rightarrow 7,2$.

Esto permitirá reflejar no solamente la adquisición de nuevas capacidades, sino también cambios de comportamiento o pérdida de consistencia respecto a capacidades previamente demostradas.

Las disminuciones requerirán, al igual que los incrementos, suficiente evidencia para evitar que una observación puntual produzca cambios artificiales.

41. Principio de no determinismo profesional

La pertenencia a un cluster profesional no supone ninguna ventaja o desventaja automática en la competición.

Los clusters se utilizarán exclusivamente para:

- Establecer el perfil inicial
- Realizar comparaciones relativas
- Calcular distancias de perfil
- Estudiar la evolución
- Analizar la transformación de capacidades

El resultado final de la competición dependerá de las reglas y métricas oficiales establecidas para cada fase.

42. Humanos, agentes de IA e inversores aleatorios

La metodología de habilidades se aplicará adaptándose a los distintos grupos participantes.

- 100.000 humanos
- 100.000 agentes de Inteligencia Artificial
- 100.000 inversores aleatorios

Todos competirán bajo las mismas reglas de inversión definidas para cada fase.

El sistema permitirá comparar no solamente su rendimiento, sino también sus patrones de comportamiento y evolución competitiva.

En el caso de los inversores aleatorios, su función principal será establecer una **referencia de comportamiento aleatorio**, permitiendo analizar qué porcentaje de estrategias humanas y de agentes de IA consigue superar sistemáticamente a dicha referencia.

43. Objetivo final de la metodología

La metodología persigue responder a dos preguntas diferentes.

Pregunta 1

¿Quién obtiene mejores resultados?

Esta pregunta será respondida mediante el Top Quant Arena y las métricas oficiales de cada fase.

Pregunta 2

¿Qué capacidades permiten obtener mejores resultados y cómo evolucionan durante la competición?

Esta pregunta será abordada mediante el Top Quant Skill Engine, los Skill Ratings, Competitive Ratings, percentiles, micro-retos, Profile Migration, Skill Growth y Learning Velocity.

De esta manera, Top Quant Arena no será únicamente una competición de resultados financieros.

La competición es también un experimento longitudinal destinado a estudiar la relación entre:

conocimiento + tecnología + razonamiento + creatividad + adaptación + comportamiento inversor + resultados.

44. Resumen del funcionamiento

El proceso completo de evaluación podrá resumirse de la siguiente manera:

1. Perfil inicial: El participante comienza con el vector de habilidades correspondiente a su cluster de origen.

↓

2. Participación: El participante toma decisiones de inversión mediante los mecanismos establecidos en cada fase.

↓

3. Observación: El sistema registra los pesos, rebalanceos, resultados, riesgos, reacciones y demás información observable.

↓

4. Micro-retos: El participante puede demostrar directamente conocimientos y capacidades específicas.

↓

5. Normalización: Las evidencias se convierten en puntuaciones comparables y percentiles.

↓

6. Competitive ELO: Se estima la capacidad del participante para superar diferentes poblaciones y percentiles.

↓

7. Skill Update: Las evidencias generan actualizaciones en las doce dimensiones de habilidades.

↓

8. Confidence: El sistema actualiza el nivel de confianza de cada estimación.

↓

9. Cluster Migration: Se analiza la distancia entre el perfil individual y los diferentes perfiles de referencia.

↓

10. Phase Ceiling: Las puntuaciones se limitan de acuerdo con las capacidades que puedan observarse razonablemente en cada fase.

↓

11. Evolución: Se calculan Skill Level, Skill Growth, Learning Velocity y Competitive Rating.

↓

12. Ranking: Los resultados alimentan tanto el Top Quant Arena como los diferentes rankings de habilidades y evolución.

45. Naturaleza de las puntuaciones

Las puntuaciones de habilidades publicadas por Top Quant Arena deberán interpretarse como estimaciones estadísticas de capacidades demostradas en el contexto de la competición.

No constituyen:

- Una certificación profesional
- Una titulación académica
- Una evaluación psicológica
- Una valoración laboral
- Una afirmación absoluta sobre las capacidades reales de una persona fuera de la competición.

Su finalidad es proporcionar una medida objetiva, dinámica y comparable de la evolución observada durante el desarrollo de Top Quant Arena.

46. Transparencia metodológica

En aras de la transparencia, AthenAI Institute of Technology hace pública la metodología general utilizada para el cálculo de los perfiles y su evolución.

- Las dimensiones evaluadas
- Los criterios generales de evidencia
- Las métricas oficiales de cada fase
- Los criterios de actualización
- Los límites máximos por fase
- Los procedimientos generales de cálculo

Los parámetros técnicos podrán ser calibrados antes del inicio de la competición mediante pruebas piloto y análisis estadísticos, con el objetivo de garantizar que el sistema sea estable, discriminante y resistente a comportamientos anómalos.

Una vez iniciada una edición, cualquier modificación sustancial de la metodología que pueda alterar materialmente la clasificación o las puntuaciones será comunicada públicamente.

47. Periodo inicial de calibración

Durante el primer mes de competición se establecerá un periodo inicial de calibración del Skill Engine.

Durante este periodo, el sistema recopilará, procesará y analizará las evidencias generadas por los participantes, incluyendo, entre otras, las decisiones de inversión, el comportamiento de las carteras, los resultados obtenidos, las respuestas a micro-retos y las demás evidencias contempladas en la presente metodología.

No obstante, **las evidencias observadas durante este primer mes no producirán actualizaciones del Skill Rating individual.** Las puntuaciones de habilidades permanecerán durante este periodo en los niveles iniciales establecidos para cada participante conforme a su perfil de origen.

El objetivo del periodo de calibración será permitir que el sistema:

- Analice la distribución inicial de las evidencias.
- Calibre los parámetros técnicos del Skill Engine.
- Ajuste, cuando resulte necesario, los procedimientos de normalización.
- Evalúe la estabilidad de las métricas utilizadas.
- Identifique posibles efectos derivados del comportamiento inicial de los participantes.
- Establezca las referencias poblacionales necesarias para las posteriores actualizaciones.
- Garantice que el proceso de actualización sea suficientemente estable, discriminante y comparable antes de comenzar a modificar los Skill Ratings.

Las evidencias generadas durante el periodo de calibración no se perderán ni serán descartadas. Permanecerán incorporadas al historial de evidencias del participante y serán utilizadas por el Skill Engine para realizar las actualizaciones posteriores, una vez finalizado dicho periodo.

Una vez concluido el primer mes, el Skill Engine comenzará a actualizar progresivamente los Skill Ratings de acuerdo con las reglas establecidas en la presente metodología.

El periodo de calibración tendrá la misma duración y se aplicará de forma homogénea a todos los participantes, sin distinción de perfil, cluster o tipo de participante.

48. Parámetros y reglas de actualización del Skill Engine

48.1. Objeto

El **Skill Engine** es el sistema encargado de transformar las evidencias observadas durante la competición en una estimación cuantitativa de las habilidades de cada participante.

Su función será actualizar semanalmente el vector de habilidades de cada participante a partir de:

- Las evidencias observadas durante la semana
- Su posición relativa respecto de la población humana
- La intensidad de dichas evidencias
- La confianza asociada a las mismas
- Su persistencia temporal
- El historial previo del participante
- El nivel de habilidad estimado anteriormente
- El Skill Ceiling correspondiente a la fase

El Skill Engine no utilizará una relación determinista del tipo:

"resultado X = aumento de Y puntos".

La actualización será el resultado de un proceso acumulativo y normalizado de inferencia.

El sistema tendrá como objetivo estimar el nivel de capacidad que los datos observados permiten atribuir razonablemente a un participante en cada dimensión de habilidad.

48.2. Escala de habilidades

Todas las habilidades se expresarán en una escala continua de 0,00 a 10,00 puntos.

Aunque los perfiles iniciales de los participantes puedan definirse mediante valores enteros, el Skill Rating evolucionará de forma continua.

Por ejemplo: 6,00 → 6,14 → 6,27 → 6,19 → 6,35

La puntuación publicada podrá mostrarse con uno o dos decimales, pero el sistema mantendrá internamente una precisión superior.

La puntuación de una habilidad estará siempre limitada por:

$$0 \leq \text{Skill} \leq \text{Skill Ceiling de la fase}$$

El Skill Ceiling representa el máximo nivel que puede ser atribuido justificadamente a una habilidad en una determinada fase.

El Skill Ceiling no representa el potencial máximo del participante, sino el máximo nivel que puede ser observado y acreditado mediante las evidencias disponibles en cada fase.

48.3. Vector de habilidades

Cada participante tendrá asociado un vector de doce habilidades:

$$S_{i,t} = [s_1, s_2, \dots, s_{12}]$$

donde:

- i identifica al participante
- t identifica la semana de evaluación
- s_k representa la puntuación de la habilidad k

Las doce dimensiones son:

1. Conocimiento Financiero
2. Conocimiento de Mercados
3. Matemáticas y Razonamiento Cuantitativo
4. Estadística y Modelización
5. Data Analytics
6. Inteligencia Artificial y Machine Learning
7. Programación
8. Diseño de Algoritmos de Inversión
9. Quant Research
10. Gestión del Riesgo
11. Implementación y Producción
12. Creatividad y Adaptación

48.4. Valor de los niveles de intensidad

Toda evidencia susceptible de modificar una habilidad tendrá asignado un nivel de intensidad.

Se utilizarán cinco niveles:

Intensidad	Valor
Muy baja	0,20
Baja	0,40
Media	0,60
Alta	0,80
Muy alta	1,00

Estos valores no representan puntos de habilidad.

Representan la capacidad potencial de una evidencia para modificar la estimación de una habilidad.

Por ejemplo, una evidencia de intensidad "Muy alta" tendrá una capacidad de contribución 2,5 veces superior a una evidencia de intensidad "Baja", antes de aplicar el resto de factores del modelo.

48.5. Valor de la confianza

Cada evidencia tendrá asociado un nivel de confianza.

Se utilizarán cuatro niveles:

Confianza	Valor
Baja	0,40
Media	0,65
Alta	0,85
Muy alta	1,00

La confianza mide hasta qué punto la evidencia permite realizar una inferencia fiable sobre una determinada habilidad.

La confianza no mide la calidad del participante.

Una evidencia puede ser extraordinariamente positiva y, sin embargo, presentar una confianza baja si se basa en una observación aislada o altamente susceptible de aleatoriedad.

48.6. Dificultad de la evidencia

Las evidencias se clasifican en cinco niveles de dificultad:

Dificultad	Valor
------------	-------

Muy baja	0,20
Baja	0,40
Media	0,60
Alta	0,80
Muy alta	1,00

La dificultad no genera directamente una subida de puntuación.

Su función es determinar la capacidad informativa de una evidencia respecto de niveles elevados de habilidad.

Una evidencia que requiere una capacidad avanzada será más informativa para distinguir entre participantes con habilidades elevadas que una evidencia que puede producirse mediante conocimientos básicos.

La dificultad podrá utilizarse especialmente en los micro-retos.

48.7. Conversión de percentiles en señales

Las métricas de rendimiento relativo se transformarán en señales normalizadas.

Para una determinada métrica, se calculará el percentil del participante dentro de la población de referencia.

El percentil se expresará como:

$$p \in [0,1]$$

donde:

- 0 representa el extremo inferior;
- 0,50 representa la mediana;
- 1 representa el extremo superior.

La señal básica será:

$$q = 2p - 1$$

Por tanto:

Percentil	Señal
0 %	-1,00
10 %	-0,80
25 %	-0,50
50 %	0,00
75 %	+0,50
90 %	+0,80
100 %	+1,00

La transformación garantiza que la mediana constituya el punto neutro.

Una posición por encima de la mediana generará una señal positiva y una posición por debajo de la mediana una señal negativa.

Cuando una métrica tenga una interpretación inversa, por ejemplo una métrica de riesgo en la que un valor elevado sea perjudicial, se invertirá previamente su orientación.

48.8. Percentiles y población de referencia

La población de referencia dependerá de la naturaleza de la evidencia.

Podrán utilizarse:

- Población humana total
- Cluster del participante
- Benchmark de mercado
- Distribución histórica del propio participante
- Distribución de una determinada fase
- Distribución de una determinada métrica

Los agentes de IA y los inversores aleatorios podrán formar parte de comparaciones específicas de la competición, pero **no formarán parte de la población utilizada para el cálculo del ELO humano**.

Cuando se utilice una comparación entre participantes humanos, la referencia estará claramente identificada.

48.9. Deciles y percentiles extremos

El sistema utilizará percentiles continuos como medida principal.

Los deciles podrán utilizarse como clasificación complementaria y para determinadas visualizaciones del ranking.

No obstante, para la actualización matemática del Skill Rating se utilizará el percentil continuo siempre que exista suficiente población estadística.

Para evitar que pequeñas diferencias entre valores extremos produzcan actualizaciones excesivas, los percentiles se limitarán internamente al intervalo [0,01 ; 0,99] antes de aplicar determinadas transformaciones estadísticas.

Por tanto, ningún participante recibirá una señal matemática equivalente a un percentil exactamente 0 % o 100 %.

48.10. ELO humano

El Skill Engine incorporará un rating ELO destinado exclusivamente a medir la capacidad competitiva relativa de los participantes humanos.

El ELO no establecerá una expectativa distinta para financieros, ingenieros, matemáticos, programadores u otros perfiles.

Todos los participantes humanos partirán de un rating inicial común:

ELO₀ = 1500

El ELO se actualizará semanalmente.

48.11. Resultado competitivo semanal

Para calcular el ELO se utilizará un **Weekly Competitive Score (WCS)**.

El WCS será una medida normalizada del rendimiento competitivo del participante durante la semana.

El WCS se construirá a partir de las métricas específicas de la fase, incluyendo cuando corresponda:

- Rendimiento
- Rendimiento ajustado por riesgo
- Estabilidad
- Drawdown
- Exposición
- Cumplimiento de las reglas y parámetros
- Otras métricas definidas en las bases de la fase

El WCS será normalizado entre los participantes humanos de la semana.

La posición relativa resultante permitirá determinar los enfrentamientos virtuales del sistema ELO.

48.12. Enfrentamientos ELO sin enfrentamiento directo

Dado que la competición no consiste en enfrentamientos individuales, el ELO no utilizará partidas directas.

Cada semana se generarán **enfrentamientos virtuales** entre participantes humanos a partir de sus resultados competitivos.

Para cada participante se seleccionarán o construirán comparaciones con participantes de distintos niveles de rating.

La selección tendrá carácter estratificado para evitar que los participantes situados en los extremos de la distribución reciban información exclusivamente de participantes muy próximos a ellos.

Como regla general, cada participante tendrá un mínimo de 32 comparaciones virtuales por semana, siempre que exista información suficiente.

Las comparaciones se distribuirán entre:

- Participantes de rating similar
- Participantes de rating inferior
- Participantes de rating superior

La selección será reproducible mediante una semilla estadística determinada por AthenAI Institute of Technology.

48.13. Resultado de un enfrentamiento virtual

Para cada comparación entre los participantes A y B se determinará un resultado:

R(A,B)

con los siguientes valores:

- **1**: A supera a B
- **0,5**: resultado equivalente o estadísticamente indistinguible
- **0**: A queda por debajo de B

El resultado se determinará mediante el WCS semanal.

Cuando la diferencia entre los WCS sea inferior a un umbral de indiferencia previamente establecido, se asignará un resultado de 0,5.

Esto evita que pequeñas diferencias atribuibles al ruido produzcan actualizaciones artificiales del rating.

48.14. Probabilidad ELO esperada

La probabilidad esperada de que A supere a B será:

$$E(A,B) = 1 / [1 + 10^{((R_B - R_A)/400)}]$$

donde:

- **R_A** es el rating ELO de A;
- **R_B** es el rating ELO de B.

Por ejemplo, si: $R_A = 1700$ y $R_B = 1500$, la probabilidad esperada de que A supere a B será aproximadamente del 76 %

Esta probabilidad representa una expectativa estadística derivada exclusivamente de los ratings previos.

48.15. Actualización del ELO

El rating se actualizará mediante:

$$R_{t+1} = R_t + K \times (O_t - E_t)$$

donde:

- R_t = rating anterior;
- K = factor de actualización;
- O_t = resultado observado;
- E_t = resultado esperado.

Se utilizará inicialmente $K = 24$ durante las primeras ocho semanas de participación.

A partir de la novena semana $K = 16$

Esta reducción permitirá que el rating se vuelva progresivamente más estable a medida que se acumula información.

Cuando se produzca un cambio de fase, el sistema podrá aplicar temporalmente un factor de adaptación para permitir que el nuevo entorno genere información nueva sin destruir completamente el historial competitivo.

48.16. Interpretación del ELO

El ELO no será interpretado como una habilidad de 0 a 10.

Su función es estimar la fuerza competitiva relativa dentro de la población humana.

A partir de la distribución de ratings podrá calcularse, entre otras métricas:

- Probabilidad estimada de superar a un participante humano seleccionado
- Probabilidad de superar a un participante humano de determinado rating
- Probabilidad de superar a un determinado percentil de participantes humanos
- Posición relativa esperada dentro de la población humana

De este modo, podrá expresarse, por ejemplo: "El rating actual del participante implica una probabilidad estimada del 90 % de superar a un participante humano situado en el rating X."

El ELO será utilizado como evidencia, no como puntuación directa de una habilidad.

48.17. ELO y clusters

Los clusters podrán utilizarse para generar poblaciones de comparación adicionales.

Se podrán calcular:

- ELO general humano
- ELO intra-cluster
- ELO inter-cluster

El ELO intra-cluster permitirá responder: ¿Cómo compite el participante frente a otros participantes cuyo comportamiento es similar?

El ELO inter-cluster permitirá analizar: ¿Cómo compite frente a participantes cuyo comportamiento presenta características diferentes?

El cambio de cluster de un participante será posible cuando sus características observadas se desplacen de forma estadísticamente significativa.

El cluster no determinará una expectativa de rendimiento.

48.18. Cálculo de la persistencia

La persistencia mide cuánto tiempo se mantiene una determinada señal.

Para cada evidencia se calculará un indicador de persistencia: $P \in [0,1]$

La persistencia se determinará a partir de la continuidad de la señal en las observaciones disponibles.

Como regla general:

Persistencia	Condición orientativa	Valor
Aislada	1 observación	0,25
Inicial	2 observaciones consecutivas	0,50
Consistente	3–4 observaciones	0,75
Persistente	≥ 5 observaciones	1,00

Cuando una evidencia requiera específicamente una duración temporal mayor, la matriz de evidencias podrá establecer un requisito de persistencia diferente.

La persistencia se reiniciará o disminuirá progresivamente cuando la señal deje de observarse.

48.19. Decaimiento temporal de la evidencia

El sistema no mantendrá indefinidamente el mismo peso de una evidencia antigua.

Cuando una evidencia deje de recibir confirmación, su peso disminuirá progresivamente.

Se utilizará un factor de decaimiento semanal: $D = 0,90$

Por tanto, una señal no confirmada mantendrá aproximadamente:

- 90 % de su peso después de una semana

- 81 % después de dos
- 72,9 % después de tres
- 65,6 % después de cuatro

Este mecanismo permite que el perfil evolucione y evita que una característica observada al comienzo de la competición determine permanentemente el perfil del participante.

48.20. Peso temporal de la evidencia

Las nuevas evidencias tendrán mayor peso que las evidencias antiguas, pero el historial no será descartado.

Para una habilidad determinada se calculará una señal histórica mediante un promedio móvil exponencial:

$$H_t = (1 - \lambda) \times H_{t-1} + \lambda \times X_t$$

donde:

- H_t = señal histórica;
- H_{t-1} = señal histórica de la semana anterior;
- X_t = señal agregada de las evidencias actuales;
- λ = factor de actualización.

Se utilizará inicialmente: $\lambda = 0,20$

Esto implica que una nueva semana tendrá un peso del 20 % en la estimación histórica y el historial acumulado conservará un peso del 80 %.

La utilización de un promedio exponencial permite que el sistema responda a cambios reales sin reaccionar excesivamente a una única observación.

48.21. Cálculo de la señal de una evidencia

Para una evidencia e que afecte a una habilidad k, se calculará:

$$X(e,k) = W(e,k) \times D(e) \times I(e,k) \times C(e,k) \times P(e) \times Q(e)$$

donde:

- $W(e,k)$ = peso estructural de la evidencia sobre la habilidad k
- $D(e)$ = dificultad
- $I(e,k)$ = intensidad
- $C(e,k)$ = confianza
- $P(e)$ = persistencia
- $Q(e)$ = señal normalizada de la observación

Los factores de intensidad, confianza, dificultad y persistencia se encuentran normalizados entre 0 y 1.

El peso estructural $\mathbf{W}(\mathbf{e}, \mathbf{k})$ será definido en la Matriz de Evidencias.

La señal $\mathbf{Q}(\mathbf{e})$ estará comprendida entre -1 y +1.

Por tanto, una evidencia positiva generará una señal positiva y una evidencia negativa una señal negativa.

48.22. Peso estructural de la evidencia

El peso estructural será específico para cada relación **Evidencia** $\times$ **Habilidad** y no únicamente para la evidencia.

Por ejemplo, una evidencia de "control persistente del drawdown" podrá tener:

- Peso muy elevado sobre Gestión del Riesgo
- Peso medio sobre Diseño de Estrategias
- Peso bajo sobre Conocimiento de Mercados
- Peso nulo sobre Programación

Esto evita asumir que una misma observación proporciona la misma información sobre todas las habilidades.

Los pesos se normalizarán internamente para evitar que una misma evidencia pueda generar una cantidad de información artificialmente superior por afectar a un mayor número de dimensiones.

48.23. Normalización de evidencias cruzadas

Cuando una misma evidencia afecte a varias habilidades, su capacidad informativa total estará limitada.

Sea: $\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_n$ el conjunto de pesos asignados a una evidencia.

Los pesos efectivos se normalizarán de forma que: $\sum |\mathbf{W}_k| \leq 1$ para cada evidencia.

De este modo, una observación no generará una cantidad de información creciente simplemente porque se haya relacionado con un mayor número de habilidades.

La información se distribuirá entre las dimensiones relevantes.

48.24. Agregación semanal

Durante cada semana podrán producirse múltiples evidencias.

Para cada habilidad se calculará una señal semanal agregada:

$$\mathbf{X}_{k,t} = \sum \mathbf{X}(\mathbf{e}, \mathbf{k}) / \sum \mathbf{A}(\mathbf{e}, \mathbf{k})$$

donde:

$$A(e,k) = W(e,k) \times I(e,k) \times C(e,k)$$

y la suma se realizará sobre las evidencias válidas de la semana.

Esta normalización evita que un participante que genere un mayor número de observaciones reciba automáticamente una mayor actualización.

Cuando no exista evidencia suficiente sobre una determinada habilidad durante una semana: $X_{k,t} = H_{k,t-1}$ La ausencia de evidencia no se interpretará como evidencia negativa.

48.25. Separación entre ausencia de evidencia y evidencia negativa

El Skill Engine distinguirá estrictamente entre:

- No observar una capacidad
- Observar un comportamiento contrario a la capacidad

La ausencia de evidencia no producirá automáticamente una reducción de la puntuación.

Por ejemplo: Si durante una fase no existen datos suficientes para evaluar Programación, la puntuación de Programación no disminuirá simplemente por no haber sido observada.

Por el contrario: Si una evidencia específica demuestra errores persistentes en programación, podrá generarse una señal negativa.

Este principio será fundamental para evitar penalizaciones artificiales.

48.26. Actualización semanal del Skill Rating

Una vez calculada la señal histórica $H_{k,t}$ se determinará el cambio semanal de la habilidad.

La fórmula general será:

$$\Delta S_{k,t} = M \times H_{k,t} \times C_t \times G_{k,t}$$

donde:

- M = máxima variación semanal;
- $H_{k,t}$ = señal histórica;
- C_t = factor global de confianza;
- $G_{k,t}$ = factor de espacio disponible respecto al Skill Ceiling o al límite inferior.

Se utilizará como máximo $M = 0,25$ puntos por semana en condiciones de máxima evidencia y confianza.

Por tanto, ninguna habilidad podrá experimentar una variación superior a **±0,25 puntos por semana** como consecuencia del proceso ordinario de actualización.

48.27. Factor de espacio respecto al Skill Ceiling

Para evitar que una habilidad alcance artificialmente el límite de fase y quede bloqueada de forma brusca, se utilizará un factor de proximidad al techo.

Si $S < \text{Ceiling}$ la distancia al techo será $U = \text{Ceiling} - S$

El sistema aplicará un factor de aproximación cuando la distancia sea inferior a 1 punto.

De forma equivalente, cuando una habilidad se encuentre próxima al Skill Ceiling, la magnitud de las actualizaciones positivas se reducirá progresivamente.

Las actualizaciones negativas no se verán limitadas de la misma forma, salvo por el límite inferior de 0.

Esto permite que alcanzar una puntuación próxima al Skill Ceiling requiera evidencias progresivamente más consistentes.

48.28. Factor de confianza global

La confianza global de una actualización será:

$$C_t = \min(1, \sqrt[N_{\text{ref}}]{N_{\text{eff}}})$$

donde:

- N_{eff} = número efectivo de observaciones independientes;
- N_{ref} = número de observaciones considerado suficiente para alcanzar confianza plena.

Se utilizará $N_{\text{ref}} = 10$

Por tanto:

Observaciones efectivas	Confianza
1	0,32
2	0,45
4	0,63
6	0,77
8	0,89
10 o más	1,00

Esto impide que una única observación extremadamente positiva o negativa produzca una modificación máxima del Skill Rating.

48.29. Observaciones efectivas

El número de observaciones efectivas no será necesariamente igual al número bruto de métricas.

Cuando varias métricas procedan del mismo fenómeno, se reducirá su contribución conjunta.

Por ejemplo:

- Rentabilidad
- Sharpe
- Volatilidad
- Drawdown

No serán consideradas cuatro observaciones completamente independientes si todas describen el mismo episodio de comportamiento. El objetivo es evitar la **dobles contabilización estadística**.

48.30. Integración del ELO en las habilidades

- El ELO no se convertirá directamente a puntos de habilidad.
- El ELO generará una evidencia relativa adicional.

La señal ELO de un participante se calculará a partir de su posición competitiva relativa y de la evolución de su rating.

Como referencia:

$Q_ELO = 2P_ELO - 1$ donde P_ELO representa la probabilidad estimada de superar al nivel de participantes utilizado como referencia.

Por ejemplo:

- Probabilidad 50 % → señal 0
- Probabilidad 75 % → señal +0,50
- Probabilidad 90 % → señal +0,80
- Probabilidad 25 % → señal -0,50

La evidencia ELO tendrá un peso específico dentro de la matriz para cada habilidad.

No todas las habilidades recibirán la misma contribución del ELO.

48.31. ELO y percentil objetivo

La metodología podrá utilizar distintos niveles de referencia.

Entre ellos:

- Mediana humana
- Percentil 75
- Percentil 90
- Percentil 95
- Determinados niveles de ELO
- Participantes pertenecientes al mismo cluster

Esto permitirá generar indicadores como: Probabilidad estimada de superar al 90 % de los participantes humanos de referencia.

La estimación será estadística y no significará que el participante haya competido directamente contra cada uno de ellos.

48.32. Actualización por cambio de fase

El cambio de fase no reiniciará el Skill Rating. El participante conservará su historial.

Sin embargo, la nueva fase podrá generar evidencias nuevas sobre habilidades que anteriormente estaban parcialmente observadas.

Cuando se produzca un cambio de fase:

1. Se conservará el Skill Rating anterior
2. Se conservará el historial de evidencias
3. Se aplicarán los nuevos Skill Ceilings
4. Se incorporarán las nuevas evidencias
5. Aumentará progresivamente la confianza sobre las habilidades que la nueva fase permita observar
6. Podrán aparecer nuevas señales de adaptación

La transición entre fases constituye en sí misma una fuente de evidencia.

48.33. Adaptación entre fases

La evolución de una habilidad no dependerá únicamente del nivel alcanzado dentro de una fase.

El sistema observará también la capacidad del participante para transferir conocimientos y adaptarse a un nuevo entorno.

La adaptación podrá generar evidencia cuando el participante:

- Mantenga un rendimiento competitivo al cambiar el universo de inversión.
- Modifique adecuadamente su comportamiento ante nuevas restricciones.
- Gestione correctamente nuevos activos financieros.
- Mantenga la robustez de sus estrategias.
- Responda adecuadamente ante nuevos tipos de riesgo.
- Demuestre capacidad para aprender y modificar su comportamiento.

La adaptación tendrá especial relevancia en las fases 3 y 4.

48.34. Cambios máximos semanales

Como regla general $|\Delta S| \leq 0,25$ para cada habilidad y semana.

No obstante, el sistema podrá producir una actualización inferior a dicho máximo en función de:

- Intensidad
- Confianza
- Persistencia
- Historial
- Dificultad
- Posición relativa
- Proximidad al Skill Ceiling

El límite semanal evita que un único evento transforme radicalmente el perfil de un participante.

48.35. Actualizaciones acumulativas

El límite semanal no implica que una habilidad solo pueda cambiar 0,25 puntos durante toda la competición. Las actualizaciones podrán acumularse.

Por ejemplo:

Semana 1: 6,00

Semana 2: 6,16

Semana 3: 6,31

Semana 4: 6,48

Semana 5: 6,62

La evolución acumulada reflejará la persistencia de las evidencias.

Por el contrario, si la evidencia positiva desaparece, el Skill Rating podrá estabilizarse o disminuir progresivamente.

48.36. Regresión hacia la evidencia reciente

El sistema no considerará que una habilidad adquirida en una etapa anterior permanece necesariamente constante durante toda la competición.

Las habilidades evolucionarán mediante la combinación de:

- Historial
- Evidencias recientes

- Persistencia
- Decaimiento temporal

Por tanto, el perfil podrá experimentar mejora, estabilidad o deterioro.

Una mejora anterior podrá perder relevancia si posteriormente aparecen evidencias persistentes que contradicen dicha estimación.

48.37. Regla de estabilidad

Para evitar oscilaciones artificiales, se aplicará una regla de estabilidad.

Una señal semanal que sea contraria a la tendencia histórica pero que no supere un nivel mínimo de intensidad y confianza tendrá un efecto limitado.

De forma general, **una única observación contradictoria no podrá revertir por sí sola una tendencia histórica consolidada.**

Para producir una reversión significativa será necesaria evidencia adicional o una señal de intensidad excepcional.

48.38. Regla de evidencia excepcional

Cuando una evidencia presente simultáneamente:

- Intensidad Muy Alta
- Confianza Muy Alta
- Dificultad Alta o Muy Alta
- Una posición relativa extrema

Podrá producir una actualización cercana al máximo semanal permitido.

Esta excepción permite que el sistema sea sensible a comportamientos realmente extraordinarios sin abandonar el principio de estabilidad.

48.39. Fórmula final del Skill Rating

La actualización general de una habilidad queda expresada de la siguiente forma:

$$S_{k,t+1} = \text{Clip}[S_{k,t} + \Delta S_{k,t} , 0 , \text{Ceiling}_{k,t}]$$

donde:

$$\Delta S_{k,t} = 0,25 \times H_{k,t} \times C_t \times G_{k,t}$$

y:

$$H_{k,t} = 0,80 \times H_{k,t-1} + 0,20 \times X_{k,t}$$

siendo:

$$X_{k,t} = \Sigma [W(e,k) \times D(e) \times I(e,k) \times C(e,k) \times P(e) \times Q(e)] / \Sigma [|W(e,k)| \times I(e,k) \times C(e,k)]$$

y:

$$Q(e) = 2p(e) - 1$$

siendo:

- $S_{k,t}$ = Skill Rating anterior
- $S_{k,t+1}$ = nuevo Skill Rating
- $Ceiling_{k,t}$ = Skill Ceiling de la fase
- $H_{k,t}$ = señal histórica
- $X_{k,t}$ = señal semanal
- $W(e,k)$ = peso de la evidencia sobre la habilidad
- $D(e)$ = dificultad
- $I(e,k)$ = intensidad
- $C(e,k)$ = confianza
- $P(e)$ = persistencia
- $Q(e)$ = señal normalizada
- $p(e)$ = percentil correspondiente
- C_t = confianza global
- $G_{k,t}$ = factor de proximidad al Skill Ceiling

La función **Clip** garantiza que $0 \leq S_{k,t+1} \leq Ceiling_{k,t}$

48.40. Ausencia de evidencia

Cuando durante una semana no exista evidencia suficiente para actualizar una habilidad:

$$S_{k,t+1} = S_{k,t}$$

La ausencia de observación no generará penalización.

El sistema podrá, no obstante, reducir progresivamente el nivel de confianza asociado a determinadas estimaciones si transcurre un periodo prolongado sin nueva información relevante.

48.41. Mínimo de información para una actualización

No se realizará una actualización significativa de una habilidad cuando la información disponible sea insuficiente para realizar una inferencia estadísticamente razonable.

Como regla general, una evidencia aislada tendrá una capacidad de actualización limitada por el factor de confianza global.

Las habilidades que no puedan ser observadas directamente en una determinada fase conservarán su última estimación válida, sujeta a las reglas de persistencia y decaimiento temporal establecidas.

48.42. Control de población

La competición podrá contar con hasta 100.000 participantes humanos.

El Skill Engine estará diseñado para trabajar sobre una población de esta magnitud sin necesidad de realizar comparaciones exhaustivas entre todos los participantes.

Las métricas poblacionales se calcularán mediante procedimientos estadísticos escalables.

Para el cálculo del ELO, los enfrentamientos virtuales se limitarán a una muestra estratificada de participantes, manteniendo simultáneamente:

- Participantes de rating similar
- Participantes de rating superior
- Participantes de rating inferior

Esto permite mantener una estimación estable sin necesidad de realizar 100.000×100.000 comparaciones cada semana.

Cuando la población sea suficientemente grande, las estimaciones de percentiles se realizarán sobre la distribución completa o mediante métodos estadísticos equivalentes que garanticen precisión suficiente.

48.43. Estabilidad estadística

Cuando una determinada métrica presente una distribución extremadamente concentrada, asimétrica o afectada por valores extremos, el sistema podrá utilizar transformaciones estadísticas antes de calcular los percentiles.

El objetivo será evitar que:

- Outliers
- Errores de medición
- Valores extremos
- Distribuciones altamente asimétricas

produzcan actualizaciones desproporcionadas.

Las transformaciones se definirán previamente para cada métrica y serán aplicadas de manera homogénea a todos los participantes.

48.44. Datos faltantes

Los datos faltantes no serán interpretados automáticamente como valores negativos.

Cuando una métrica no pueda calcularse de manera fiable:

1. No se generará una evidencia negativa
2. Se reducirá, cuando corresponda, la confianza de la observación
3. La habilidad podrá permanecer sin cambios
4. El sistema esperará a disponer de información suficiente

48.45. Control frente al ruido

El Skill Engine incorporará mecanismos destinados a diferenciar señal de ruido.

Entre ellos:

- Normalización por percentiles
- Persistencia temporal
- Decaimiento de evidencias antiguas
- Límites máximos de variación semanal
- Confianza estadística
- Agregación de observaciones
- Control de doble contabilización
- Comparación relativa
- Estabilidad del historial

El objetivo es que el Skill Rating represente una estimación de capacidad y no una reacción inmediata a las fluctuaciones aleatorias del mercado.

48.46. Control frente al sobreajuste del perfil

El sistema no deberá modificar significativamente una habilidad cuando la evidencia disponible pueda explicarse razonablemente por variación aleatoria.

Por esta razón, una observación extremadamente favorable no será suficiente por sí sola para transformar sustancialmente el perfil de un participante.

La evolución significativa requerirá, como regla general:

magnitud + consistencia + información suficiente.

Del mismo modo, un periodo adverso no implicará automáticamente un deterioro permanente de las habilidades.

48.47. Reproducibilidad

Todas las reglas del Skill Engine serán aplicadas de manera homogénea a todos los participantes.

Para garantizar la reproducibilidad:

- Los parámetros están definidos previamente
- Las métricas están definidas previamente
- Las poblaciones de referencia están determinadas por reglas objetivas
- Los factores de intensidad y confianza están normalizados
- Los Skill Ceilings están establecidos por fase
- Los criterios de actualización son comunes
- Cualquier proceso de muestreo estadístico utilizará procedimientos reproducibles

No se realizarán ajustes individuales de las puntuaciones.

48.48. Principio final del Skill Engine

El Skill Engine no pretende responder a la pregunta:

"¿Cuánto dinero ha ganado el participante?"

Su objetivo es responder a una pregunta diferente:

"¿Qué evidencias ha generado el participante durante la competición que permitan estimar qué capacidades posee, con qué nivel de confianza y cómo están evolucionando?"

Por ello, el resultado financiero constituye únicamente una de las posibles fuentes de información.

El perfil final será consecuencia de la acumulación de evidencias procedentes de:

- Decisiones de inversión
- Comportamiento de las carteras
- Gestión del riesgo
- Comportamiento ante distintos mercados
- Robustez de las estrategias
- Rendimiento relativo
- ELO frente a participantes humanos
- Micro-retos
- Adaptación entre fases
- Persistencia de los comportamientos

El sistema permitirá, de esta manera, que un participante pueda mejorar sustancialmente determinadas habilidades, permanecer estable en otras y experimentar un deterioro en algunas de ellas.

El **perfil** no será estático. Será una **estimación dinámica de las capacidades demostradas durante la competición.**

La puntuación final de cada participante será, por tanto, el resultado de la interacción entre: **Evidencia + comparación relativa + intensidad + confianza + persistencia + historial + dificultad + fase** y no de una única métrica de rendimiento.

48.49. **Peso estructural $W(e,k)$ de la Matriz de Evidencias**

El **peso estructural $W(e,k)$** determina la intensidad y dirección de la relación entre una evidencia e y una habilidad k .

Los valores posibles serán:

Valor	Significado
+1,00	Relación directa y muy fuerte
+0,75	Relación directa y fuerte
+0,50	Relación relevante
+0,25	Relación indirecta o secundaria
0,00	Sin relación suficiente
-0,25	Relación negativa secundaria
-0,50	Relación negativa relevante
-0,75	Relación negativa fuerte
-1,00	Relación negativa directa y muy fuerte

El signo determina la dirección del efecto y el valor absoluto determina la intensidad de la relación estructural. $W(e,k)$ no representa por sí mismo una variación del Skill Rating, sino exclusivamente la capacidad de la evidencia para informar sobre la habilidad correspondiente. El cambio efectivo de la puntuación se determinará posteriormente mediante la intensidad, dificultad, confianza, persistencia, percentil, independencia de la evidencia, fase y demás parámetros definidos en el Skill Engine.

La siguiente matriz constituye la **Matriz de Pesos Estructurales** utilizada por el Skill Engine. Las filas representan evidencias y las columnas habilidades.

Evidencia e	Financiero	Mercados	Matemáticas	Estadística	Data	AI/ML	Programación	Estrategias	Quant Research	Riesgo	Producción	Creatividad /Adaptación
Valoración correcta de activos	1,00	0,50	0,75	0,25	0,25	0,00	0,00	0,50	0,50	0,25	0,00	0,25
Construcción eficiente de cartera	0,75	0,50	0,50	0,50	0,25	0,00	0,00	1,00	0,50	0,75	0,25	0,25
Selección de activos	0,75	0,75	0,25	0,50	0,25	0,00	0,00	1,00	0,50	0,50	0,00	0,50
Diversificación eficiente	0,50	0,50	0,50	0,50	0,00	0,00	0,00	0,75	0,25	1,00	0,00	0,25
Formación de precios	0,50	1,00	0,50	0,25	0,00	0,00	0,00	0,50	0,50	0,25	0,00	0,50
Gestión de liquidez	0,25	1,00	0,25	0,25	0,00	0,00	0,00	0,50	0,25	0,75	0,50	0,25
Gestión de volatilidad	0,25	0,75	0,50	0,75	0,00	0,00	0,00	0,25	0,50	1,00	0,25	0,25
Análisis de correlaciones	0,50	0,50	0,75	0,75	0,25	0,00	0,00	0,50	0,75	0,75	0,00	0,25
Ejecución eficiente	0,25	1,00	0,25	0,25	0,00	0,00	0,25	0,50	0,25	0,50	1,00	0,25
Control del impacto de mercado	0,25	1,00	0,50	0,50	0,00	0,00	0,25	0,50	0,50	0,75	0,75	0,25
Resultado relativo frente al universo	0,50	0,50	0,25	0,50	0,25	0,25	0,00	0,75	0,75	0,50	0,25	0,50
Estabilidad del resultado	0,50	0,50	0,25	0,75	0,25	0,25	0,00	0,50	1,00	0,75	0,25	0,50
Robustez entre periodos	0,25	0,50	0,50	0,75	0,25	0,25	0,00	0,75	1,00	0,75	0,25	0,75
Robustez ante cambios de parámetros	0,25	0,25	0,75	0,75	0,25	0,50	0,25	0,75	1,00	0,50	0,25	0,75
Control del overfitting	0,00	0,25	0,75	0,75	0,50	0,50	0,25	0,75	1,00	0,50	0,25	0,75

Evidencia e	Financiero	Mercados	Matemáticas	Estadística	Data	AI/ML	Programación	Estrategias	Quant Research	Riesgo	Producción	Creatividad /Adaptación
Validación fuera de muestra	0,25	0,25	0,50	1,00	0,50	0,50	0,25	0,75	1,00	0,50	0,25	0,50
Calidad y consistencia de datos	0,00	0,25	0,25	0,50	1,00	0,50	0,50	0,25	0,50	0,25	0,50	0,25
Uso eficiente de información	0,25	0,50	0,25	0,50	1,00	0,50	0,25	0,75	0,75	0,25	0,25	0,75
Capacidad predictiva	0,25	0,50	0,75	1,00	0,75	0,75	0,25	0,75	0,75	0,50	0,25	0,50
Eficiencia de modelo ML	0,00	0,25	0,50	0,75	0,50	1,00	0,50	0,50	0,75	0,25	0,50	0,50
Generalización de modelos	0,00	0,25	0,50	0,75	0,50	1,00	0,50	0,50	1,00	0,50	0,50	0,75
Uso de modelos generativos/agénticos	0,00	0,00	0,25	0,25	0,50	1,00	0,75	0,75	0,50	0,25	0,75	1,00
Eficiencia computacional	0,00	0,25	0,50	0,25	0,50	0,75	1,00	0,25	0,50	0,00	1,00	0,50
Calidad de implementación	0,00	0,25	0,25	0,25	0,50	0,50	1,00	0,50	0,50	0,25	1,00	0,50
Automatización de procesos	0,00	0,25	0,00	0,00	0,50	0,50	1,00	0,50	0,25	0,25	1,00	0,50
Diseño de señales	0,50	0,50	0,50	0,50	0,25	0,50	0,25	1,00	0,75	0,50	0,25	0,75
Selección de variables	0,25	0,50	0,50	0,75	0,75	0,50	0,25	0,75	1,00	0,25	0,00	0,75
Asignación eficiente de capital	0,75	0,50	0,75	0,50	0,00	0,00	0,00	1,00	0,75	1,00	0,25	0,25
Control de drawdown	0,25	0,50	0,25	0,50	0,00	0,00	0,00	0,50	0,50	1,00	0,25	0,50
Control de VaR y pérdidas extremas	0,50	0,50	0,50	0,75	0,00	0,00	0,00	0,25	0,50	1,00	0,25	0,25
Control de exposición	0,50	0,50	0,25	0,25	0,00	0,00	0,00	0,50	0,25	1,00	0,25	0,25
Position sizing eficiente	0,50	0,25	0,75	0,50	0,00	0,00	0,00	0,75	0,50	1,00	0,00	0,25
Stress testing	0,25	0,50	0,50	0,75	0,00	0,25	0,00	0,50	0,75	1,00	0,00	0,75
Gestión persistente del riesgo	0,25	0,50	0,25	0,50	0,00	0,00	0,00	0,50	0,50	1,00	0,25	0,75
Exceso de turnover sin mejora de resultado	-0,25	-0,50	0,00	0,00	0,00	0,00	0,00	-0,50	-0,75	-0,50	-0,25	-0,25
Drawdown excesivo no controlado	-0,25	-0,25	0,00	-0,25	0,00	0,00	0,00	-0,50	-0,50	-1,00	-0,25	-0,50
Inestabilidad entre periodos	-0,25	-0,25	0,00	-0,50	0,00	0,00	0,00	-0,50	-1,00	-0,50	-0,25	-0,75
Sensibilidad excesiva a parámetros	0,00	0,00	-0,50	-0,50	-0,25	-0,50	-0,25	-0,75	-1,00	-0,25	-0,25	-0,75
Degradación significativa fuera de muestra	-0,25	-0,25	-0,50	-0,75	-0,25	-0,50	-0,25	-0,75	-1,00	-0,50	-0,25	-0,75
Reacción adecuada ante cambio de régimen	0,50	0,75	0,50	0,50	0,25	0,25	0,00	0,75	0,75	0,50	0,25	1,00
Adaptación a un nuevo mercado o instrumento	0,50	0,50	0,50	0,50	0,50	0,50	0,25	0,75	0,75	0,50	0,50	1,00
Transferencia de conocimiento entre fases	0,50	0,50	0,50	0,50	0,50	0,50	0,25	0,75	1,00	0,50	0,50	1,00
Descubrimiento de nuevas relaciones	0,25	0,50	0,50	0,75	0,75	0,75	0,25	0,75	1,00	0,25	0,25	1,00
Solución no convencional con resultado robusto	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,75	0,75	0,50	0,50	1,00

Los pesos anteriores constituyen los valores estructurales de referencia del Skill Engine. No se modificarán en función del resultado obtenido por un participante concreto ni se recalibrarán retrospectivamente para explicar los resultados de la competición.

Cuando una evidencia presente relación con varias habilidades, cada relación se evaluará de forma independiente mediante el correspondiente valor $W(e,k)$. El efecto final sobre cada habilidad dependerá adicionalmente de los restantes parámetros del Skill Engine.

Los valores de $W(e,k)$ serán comunes a todos los participantes y a todas las fases. La distinta capacidad de cada fase para observar determinadas habilidades se determinará mediante los Skill Ceilings y los factores de observabilidad definidos en el resto de la metodología.

48.49. Definición exacta del catálogo de evidencias e

El Skill Engine utilizará un catálogo cerrado de evidencias observables. Cada evidencia será una variable calculada exclusivamente a partir de los datos que la plataforma reciba de las carteras, de sus operaciones, de sus rebalanceos, de los activos disponibles, de las condiciones de mercado y de los resultados registrados durante la competición.

El sistema no utilizará como evidencia el código fuente, la arquitectura interna, los modelos, los algoritmos, las instrucciones proporcionadas a agentes de IA ni ninguna otra información no observable directamente por la plataforma.

Cada evidencia será identificada mediante un código único y tendrá asociada una definición matemática, una unidad de medida y las dimensiones de habilidad sobre las que puede actuar mediante la Matriz $W(e,k)$.

El catálogo de evidencias es el siguiente:

Código	Evidencia	Descripción
E01	Rentabilidad relativa	Rendimiento de la cartera respecto del universo de participantes durante el periodo de observación
E02	Rentabilidad ajustada por riesgo	Rendimiento obtenido en relación con el riesgo asumido
E03	Consistencia del resultado	Estabilidad del rendimiento durante los distintos periodos de observación
E04	Posición relativa en el ranking	Posición del participante respecto del conjunto comparable de participantes
E05	Percentil de rendimiento	Percentil ocupado por el participante en la distribución de resultados
E06	Persistencia del rendimiento	Capacidad de mantener una posición relativa elevada durante sucesivos periodos
E07	Volatilidad de cartera	Volatilidad de los rendimientos de la cartera
E08	Drawdown máximo	Máxima pérdida acumulada desde un máximo histórico de la cartera
E09	Control del drawdown	Capacidad de limitar y recuperar drawdowns en relación con el universo
E10	VaR empírico	Pérdida correspondiente a un determinado percentil de la distribución de rendimientos
E11	Pérdida extrema	Magnitud de las pérdidas situadas en la cola de la distribución
E12	Exposición	Exposición de la cartera a los distintos activos, sectores, factores o posiciones disponibles
E13	Concentración	Grado de concentración de la cartera
E14	Diversificación	Grado de diversificación efectiva de la cartera
E15	Estabilidad de exposición	Variación de las exposiciones a lo largo del tiempo
E16	Turnover	Volumen de rotación de la cartera
E17	Turnover ajustado por resultado	Relación entre rotación adicional y mejora obtenida en rendimiento o riesgo
E18	Coste de transacción relativo	Impacto de los costes de ejecución sobre el resultado
E19	Impacto de mercado	Pérdida de rendimiento atribuible al impacto de las operaciones sobre el mercado, cuando pueda estimarse
E20	Eficiencia de ejecución	Diferencia entre el precio efectivo de ejecución y el precio de referencia disponible
E21	Timing de rebalanceo	Calidad relativa de las decisiones de rebalanceo respecto de las condiciones posteriores de mercado

E22	Estabilidad entre periodos	Variación del comportamiento de la cartera entre distintos periodos de observación
E23	Robustez ante cambios de mercado	Estabilidad del comportamiento ante diferentes condiciones o regímenes de mercado
E24	Robustez fuera de muestra	Diferencia entre el comportamiento observado en periodos de referencia y periodos posteriores, cuando exista separación temporal válida
E25	Sensibilidad a cambios de parámetros observables	Variación del comportamiento cuando puedan realizarse perturbaciones estandarizadas sobre los parámetros de inversión permitidos
E26	Degradación fuera de muestra	Pérdida relativa de rendimiento o calidad al pasar de un periodo de referencia a un periodo no utilizado para su construcción, cuando dicha información pueda determinarse
E27	Correlación con el mercado	Dependencia estadística entre el rendimiento de la cartera y el rendimiento del mercado de referencia
E28	Correlación con otros participantes	Grado de similitud estadística entre la evolución de la cartera y la del resto de participantes
E29	Diversificación temporal	Capacidad de mantener comportamiento favorable en diferentes ventanas temporales
E30	Adaptación a cambios de régimen	Variación del comportamiento antes y después de cambios objetivamente identificables en las condiciones de mercado
E31	Adaptación entre fases	Capacidad de mantener o mejorar el comportamiento al cambiar el universo de inversión y las reglas de la competición
E32	Transferencia de rendimiento	Conservación de la ventaja relativa obtenida al pasar a un entorno de inversión diferente
E33	Velocidad de adaptación	Tiempo necesario para recuperar un comportamiento competitivo después de un cambio de entorno
E34	Eficiencia de capital	Rendimiento obtenido por unidad de capital o riesgo utilizado
E35	Eficiencia de diversificación	Mejora de la relación rentabilidad-riesgo asociada a la diversificación efectivamente utilizada
E36	Estabilidad de la estrategia observable	Grado de estabilidad del comportamiento de las posiciones y exposiciones a lo largo del tiempo
E37	Capacidad predictiva observable	Capacidad de determinadas decisiones de cartera para anticipar movimientos posteriores del mercado
E38	Calidad de información observable	Calidad, completitud y consistencia de los datos utilizados cuando el participante deba seleccionar información disponible en la plataforma
E39	Selección eficiente de activos	Diferencia entre el comportamiento de los activos seleccionados y alternativas comparables disponibles
E40	Asignación eficiente de capital	Calidad de la distribución del capital entre activos en términos de rendimiento y riesgo
E41	Solución no convencional con resultado robusto	Existencia de un comportamiento significativamente diferenciado del patrón predominante y simultáneamente robusto
E42	Descubrimiento de relaciones	Capacidad observable para obtener ventajas asociadas a relaciones entre activos, variables o mercados que no sean comunes al universo de participantes
E43	Cambio de estrategia	Modificación observable de las exposiciones ante cambios relevantes de mercado
E44	Control de pérdidas	Capacidad para reducir exposición después de episodios adversos y limitar pérdidas posteriores
E45	Recuperación del drawdown	Velocidad relativa de recuperación después de una pérdida acumulada
E46	Asimetría rentabilidad-riesgo	Relación entre participación en ganancias y participación en pérdidas
E47	Estabilidad de las correlaciones	Variación temporal de las correlaciones entre las posiciones de la cartera
E48	Exceso de turnover sin mejora	Rotación adicional que no genera una mejora estadísticamente significativa de rendimiento ajustado por riesgo
E49	Drawdown excesivo no controlado	Episodios de pérdida acumulada significativamente superiores a los niveles comparables del universo

E50	Inestabilidad entre periodos	Deterioro significativo y persistente del comportamiento entre periodos consecutivos
E51	Sensibilidad excesiva	Pérdida significativa de calidad ante pequeñas perturbaciones de las condiciones observables
E52	Reacción adecuada ante cambio de régimen	Mejora o preservación significativa de la posición relativa después de un cambio de régimen
E53	Adaptación a nuevo instrumento	Capacidad de obtener resultados competitivos tras la introducción de un nuevo tipo de activo o instrumento
E54	Adaptación a nueva información	Capacidad de modificar las exposiciones de forma adecuada después de la publicación de nueva información disponible para todos los participantes

Las evidencias E48-E51 constituyen evidencias de carácter potencialmente negativo. Su signo será determinado por la definición de la evidencia y por la Matriz W(e,k).

El catálogo será común a los participantes. No se crearán evidencias específicas para individuos concretos.

Una evidencia no se considerará disponible cuando no exista información suficiente para calcularla con un nivel mínimo de fiabilidad. En ese caso, su contribución al Skill Engine será igual a cero durante el periodo correspondiente y no se interpretará la ausencia de evidencia como evidencia positiva ni negativa.

Las evidencias podrán ser utilizadas en una o varias fases. La capacidad de cada fase para generar información sobre una evidencia se determinará mediante los parámetros de observabilidad y los Skill Ceilings definidos en la metodología.

El catálogo podrá incorporar evidencias adicionales cuando una nueva fase introduzca información que no pueda ser representada adecuadamente por las evidencias existentes. Cualquier nueva evidencia deberá definirse matemáticamente antes de ser utilizada y deberá incorporarse a la metodología de forma homogénea para todos los participantes.

48.50. Cálculo de las evidencias a partir de los datos de la competición

Todas las evidencias se calcularán a partir de una serie temporal de valoración de la cartera y de los registros de posición y rebalanceo proporcionados por la plataforma.

Para cada participante i y periodo de observación t , se construirá una serie de valores de cartera $V(i,t)$ y una serie de rendimientos:

$$r(i,t) = V(i,t) / V(i,t-1) - 1$$

Cuando proceda, los rendimientos se ajustarán por costes de transacción, comisiones, slippage y demás costes que formen parte de las reglas de la fase.

El periodo estándar de evaluación será semanal. Cuando una evidencia requiera una ventana temporal superior, se utilizará una ventana móvil de las últimas semanas disponibles. La ventana concreta será la definida para cada evidencia en la configuración técnica de la competición.

Las evidencias se calcularán de la siguiente forma:

Código	Evidencia	Cálculo
--------	-----------	---------

E01	Rentabilidad relativa	$R_i = \Pi(1+r_{i,t}) - 1$; posteriormente se calculará su posición relativa frente al universo
E02	Rentabilidad ajustada por riesgo	R_i / σ_i , utilizando la volatilidad calculada sobre la misma ventana
E03	Consistencia del resultado	$1 - CV(R_{\text{weekly}})$ normalizado respecto del universo
E04	Posición relativa	rank_i / N , invirtiendo la escala para que el mejor participante obtenga 1
E05	Percentil de rendimiento	Percentile(R_i) dentro del universo humano comparable
E06	Persistencia del rendimiento	Media móvil de los percentiles semanales de las últimas m semanas
E07	Volatilidad	$\sigma_i = \text{std}(r_{i,t}) \times \sqrt{A}$, donde A es el factor de anualización definido para la frecuencia
E08	Drawdown máximo	$MDD_i = \max_t [(\text{Peak}_i(t) - V_i(t)) / \text{Peak}_i(t)]$
E09	Control del drawdown	Comparación del MDD y de la duración de los drawdowns del participante con la distribución del universo
E10	VaR empírico	Cuantil q de la distribución de rendimientos: $\text{VaR}_q = -\text{Quantile}_q(r_i)$
E11	Pérdida extrema	Media de los peores q% rendimientos de la cartera
E12	Exposición	' $\text{Exposure}_{j,t} = w_{j,t}$ '; se evaluarán exposición máxima, media y variación
E13	Concentración	Índice Herfindahl-Hirschman: $\text{HHI} = \sum_j w_{j,t}^2$
E14	Diversificación	$1 - \text{HHI}$, normalizado al universo de activos disponible
E15	Estabilidad de exposición	' $1 - \text{mean}(w_{j,t} - w_{j,t-1})$ ' normalizado
E16	Turnover	' $\text{Turnover}_t = 0,5 \times \sum_j w_{j,t} - w_{j,t-1} $
E17	Turnover ajustado por resultado	Mejora de rendimiento ajustado por riesgo dividida por turnover adicional
E18	Coste de transacción relativo	$\text{Costes}_i / \text{Capital}_i$ durante el periodo
E19	Impacto de mercado	Diferencia entre el precio de referencia y el precio efectivo de ejecución, cuando exista información suficiente
E20	Eficiencia de ejecución	' $\text{Execution}_i = 1 - P_{\text{execution}} - P_{\text{reference}} / P_{\text{reference}}$ ', normalizado
E21	Timing de rebalanceo	Rendimiento posterior de las variaciones de peso respecto de las alternativas disponibles en el momento del rebalanceo
E22	Estabilidad entre periodos	' $1 - \text{Metric}_t - \text{Metric}_{t-1}$ ', normalizado
E23	Robustez ante mercado	Comparación del percentil del participante entre diferentes regímenes de mercado
E24	Robustez fuera de muestra	Diferencia entre métrica del periodo de referencia y métrica del periodo posterior
E25	Sensibilidad a parámetros	Variación de la métrica ante perturbaciones estandarizadas definidas por la competición
E26	Degradación fuera de muestra	$\text{Degradation} = \text{Metric}_{\text{IS}} - \text{Metric}_{\text{OOS}}$, normalizada por la desviación del universo
E27	Correlación con mercado	$\rho_{i,M} = \text{Corr}(r_i, r_M)$
E28	Correlación con participantes	Media o mediana de $\text{Corr}(r_i, r_j)$ respecto del universo humano
E29	Diversificación temporal	Media de los percentiles obtenidos en distintas ventanas temporales
E30	Adaptación a régimen	Diferencia entre el percentil antes y después del cambio de régimen
E31	Adaptación entre fases	Percentil obtenido durante las primeras semanas de la nueva fase comparado con el percentil final de la fase anterior
E32	Transferencia de rendimiento	Persistencia de la ventaja relativa al cambiar el universo de inversión
E33	Velocidad de adaptación	Número de semanas necesarias para alcanzar un percentil objetivo después del cambio de fase

E34	Eficiencia de capital	Return / Risk o métrica equivalente definida para la fase
E35	Eficiencia de diversificación	Mejora del ratio rentabilidad-riesgo respecto de una cartera equivalente sin la diversificación observada
E36	Estabilidad de estrategia observable	$1 - \text{mean}(\Delta w_j)$ normalizado
E37	Capacidad predictiva observable	Correlación entre cambios de posición y rendimientos posteriores de los activos afectados
E38	Calidad de información	Proporción de datos válidos, completos y consistentes cuando la fase permita observar la selección de información
E39	Selección eficiente	Rendimiento posterior de los activos seleccionados frente al conjunto de activos elegibles
E40	Asignación eficiente	Diferencia entre el rendimiento ajustado por riesgo obtenido y el de una asignación de referencia
E41	Solución no convencional robusta	Diferencia entre el patrón de exposición del participante y el patrón predominante del universo, condicionada a persistencia de la ventaja relativa
E42	Descubrimiento de relaciones	Ventaja relativa asociada a relaciones de rendimiento/correlación que presenten baja frecuencia en el universo
E43	Cambio de estrategia	Magnitud de cambio en la distribución de pesos después de un cambio significativo de mercado
E44	Control de pérdidas	Reducción de exposición después de pérdidas y comparación con la pérdida posterior evitada
E45	Recuperación del drawdown	Tiempo necesario para recuperar el máximo histórico, comparado con el universo
E46	Asimetría	Upside Capture / Downside Capture, calculados frente al mercado de referencia
E47	Estabilidad de correlaciones	$1 - \text{std}(\rho_t)$ normalizado
E48	Exceso de turnover	Turnover superior al percentil de referencia sin mejora estadísticamente significativa de E02
E49	Drawdown excesivo	MDD situado en los percentiles extremos de la distribución y acompañado de ausencia de recuperación
E50	Inestabilidad	Deterioro persistente del percentil durante periodos consecutivos
E51	Sensibilidad excesiva	Variación de la métrica superior al umbral establecido ante perturbaciones pequeñas
E52	Reacción adecuada	Cambio de percentil posterior a un cambio de régimen superior al esperado por la distribución
E53	Adaptación a nuevo instrumento	Percentil obtenido durante las primeras m observaciones del nuevo instrumento
E54	Adaptación a nueva información	Diferencia entre la respuesta de la cartera y la respuesta media del universo después de un evento informativo común

Las evidencias cuya definición dependa de una comparación entre participantes se calcularán exclusivamente respecto del universo humano elegible para la comparación. Los agentes de IA y los inversores aleatorios no formarán parte del universo de referencia utilizado para actualizar el Skill Rating de los humanos.

Para evitar que las diferencias de escala entre evidencias distorsionen el Skill Engine, cada evidencia será transformada antes de su utilización en una señal normalizada.

Para las evidencias en las que un valor superior represente un mejor comportamiento:

$$Q(e,i,t) = 2 \times \text{Percentile}(e,i,t) - 1$$

De esta forma:

- Percentil 0 $\rightarrow Q = -1$
- Percentil 0,25 $\rightarrow Q = -0,50$

- Percentil 0,50 $\rightarrow Q = 0$
- Percentil 0,75 $\rightarrow Q = +0,50$
- Percentil 1 $\rightarrow Q = +1$

Para las evidencias en las que un valor inferior represente un mejor comportamiento, se utilizará:

$$Q(e,i,t) = 1 - 2 \times \text{Percentile}(e,i,t)$$

La señal Q será, por tanto, siempre comprendida entre -1 y $+1$.

Cuando una evidencia tenga una naturaleza negativa, por ejemplo, exceso de turnover, drawdown excesivo o degradación fuera de muestra, se utilizará directamente la transformación inversa, de manera que un comportamiento favorable genere una señal positiva y un comportamiento desfavorable una señal negativa.

Los percentiles se calcularán utilizando métodos robustos frente a empates. Cuando existan más de 100.000 observaciones, se utilizará el percentil empírico exacto o, cuando resulte computacionalmente necesario, un método equivalente de estimación que preserve el orden relativo de los participantes.

Para evitar que valores extremos individuales distorsionen las distribuciones, las métricas continuas podrán someterse a winsorización (reemplazo de valores atípicos) en los percentiles 1 y 99 antes del cálculo del percentil, siempre que dicha transformación esté definida previamente para la evidencia correspondiente.

Las evidencias que requieran un número mínimo de observaciones no serán activadas hasta alcanzar dicho mínimo. La ausencia de observaciones suficientes generará $Q(e,i,t) = 0$, pero no se interpretará como un rendimiento situado en el percentil 50. El valor cero en este caso significa evidencia no disponible, y su peso será excluido del denominador de la agregación semanal.

Para cada evidencia se calculará asimismo un nivel de confianza $C(e,i,t)$ en función de:

- Número de observaciones disponibles
- Longitud de la ventana
- Estabilidad de la métrica
- Presencia de datos faltantes
- Dependencia temporal
- Robustez estadística de la estimación

La confianza será normalizada en el intervalo $[0,1]$.

Cuando una evidencia dependa de una comparación relativa con otros participantes, el universo de comparación estará constituido por los participantes humanos que hayan cumplido las condiciones mínimas de participación durante el periodo correspondiente.

Los resultados de los agentes de IA y de las estrategias aleatorias se mostrarán separadamente en los rankings específicos de la competición, pero no modificarán el percentil utilizado para actualizar las habilidades de los participantes humanos.

La información procedente de las diferentes evidencias será agregada posteriormente mediante la matriz $W(e,k)$ y los restantes parámetros definidos en el Skill Engine.

En consecuencia, el flujo de cálculo será:

Datos de cartera $\rightarrow$ Métrica observable $\rightarrow$ Evidencia $E(e) \rightarrow$ Percentil $\rightarrow$ Señal $Q(e) \rightarrow$ Confianza $C(e) \rightarrow$ Peso estructural $W(e,k) \rightarrow$ Intensidad/Dificultad/Persistencia $\rightarrow$ Agregación semanal $\rightarrow$ Actualización del Skill Rating.

Ninguna evidencia podrá modificar directamente una habilidad sin pasar por este proceso.