

## Tema: Reflexión sobre el futuro de la IA y sus implicaciones éticas y sociales

---

### ESTRUCTURA DE LA PRESENTACIÓN (5 minutos)

#### PERSONA 1 - Introducción, Contexto y Proceso de Trabajo (1:40 min)

**[Inicio]** "Boas. Hoy vamos a hablar sobre el futuro de la Inteligencia Artificial basándonos en un artículo de Senén Barro Ameneiro, director del CiTIUS.

**[Metodología - CRÍTICA]** Pero antes de nada, queremos ser transparentes sobre nuestro proceso de trabajo. Usamos ChatGPT para ayudarnos a estructurar esta presentación, y esto nos llevó a una reflexión irónica: estamos usando IA para hablar sobre los peligros de la IA.

**[Análisis crítico del proceso]** Detectamos varios problemas: - ChatGPT tendió a darnos **respuestas excesivamente optimistas** cuando le pedimos soluciones - Tuvo que ser **reorientado varias veces** porque generaba contenido genérico - **No captó inicialmente** la profundidad filosófica del artículo de Barro sobre la deshumanización

Esto nos hizo debatir entre nosotros: ¿hasta qué punto podemos confiar en que una IA analice críticamente... a la propia IA? Decidimos que el modelo es útil como **asistente de estructura**, pero el **análisis crítico tiene que ser nuestro**.

**[Conexión con Cajal]** El punto de partida es fascinante: Santiago Ramón y Cajal escribió en 1906 un relato titulado 'La vida en el año 6000'. Imaginó telemedicina, manipulación genética, máquinas diagnosticando con matemáticas complejas... Lo extraordinario: pensaba que necesitaríamos 4.000 años. Estamos aquí en 2025 con muchas de esas tecnologías ya funcionando.

**[Transición]** Si Cajal se equivocó tanto en los plazos... ¿qué nos deparará realmente la IA?"

---

#### PERSONA 2 - Riesgos, Distopías y Debate con IA (1:40 min)

**[Toma el relevo]** "Y aquí viene lo inquietante. Senén Barro reconoce que cuando imaginamos el futuro de la IA, casi siempre acabamos en distopías terroríficas.

**[Reflexión sobre el debate con ChatGPT]** Cuando le preguntamos a ChatGPT sobre estos riesgos, algo curioso pasó: nos dio listas perfectamente organizadas de peligros, pero sin apenas **peso emocional o urgencia**. Era como leer un manual técnico sobre el apocalipsis.

Le preguntamos: "*¿Crees que la IA podría acabar con la humanidad?*" Su respuesta fue diplomática y llena de 'depende', 'podría', 'es importante que'... Nos dio una **alucinación de control**: la sensación de que todo está bajo control si seguimos 'buenas prácticas'.

**[Nuestra discrepancia con el modelo]** Nosotros tres debatimos y concluimos algo diferente: **no tenemos tanto control como ChatGPT sugiere**. Los riesgos son más urgentes:

- **Armas autónomas**: ya existen prototipos, no son ciencia ficción
- **Suplantación de identidad**: deepfakes cada vez más indetectables
- **Desigualdad brutal**: las grandes tecnológicas concentran el poder de la IA
- **Deshumanización sutil**: delegamos pensamiento crítico sin darnos cuenta

**[Punto clave - coincidencia con Barro]** Aquí sí coincidimos con el artículo: el relato de Cajal sobre eliminar el amor con una vacuna no es tan absurdo. Ya estamos **externalizando capacidades** que nos hacen humanos: memoria, navegación, escritura, incluso conversación. ¿Dónde está el límite?"

---

### **PERSONA 3 - Visión Esperanzadora, Discrepancias y Conclusión (1:40 min)**

**[Cambio de tono - esperanzador pero realista]** "Ahora bien, no queremos caer en el catastrofismo. Senén Barro nos ofrece una visión de cómo DEBERÍA ser la IA del futuro.

**[Análisis crítico de nuestra conversación con ChatGPT]** Le pedimos al modelo que nos describiera una IA ética. Nos dio cuatro principios hermosos: útil, ubicua pero silente, potenciadora y democrática. Todo perfecto... **demasiado perfecto**.

**[Nuestra reflexión crítica]** Entre los tres debatimos y encontramos **contradicciones que el modelo ignoró**:

1. **"Útil pero no dañina"** - ¿Quién decide qué es útil? Un dron puede ser útil para repartir medicinas... o para vigilancia masiva. ChatGPT no profundizó en esto.
2. **"Ubicua pero silente"** - Hay una tensión: cuanto más ubicua, menos silente. Ya vivimos con algoritmos que deciden qué vemos, leemos, compramos... y NO son silentes, **manipulan nuestra atención**.
3. **"Democrática y accesible"** - Le preguntamos: "*¿Cómo se hace la IA democrática si cuesta millones entrenar un modelo?*" Su respuesta fue vaga: 'iniciativas de código abierto', 'colaboración internacional'... **No reconoció que el modelo económico actual va en dirección opuesta**.

**[Punto de acuerdo]** Coincidimos con Barro en la urgencia: **no podemos esperar al año 6000 para decidir**. Pero aquí descubrimos una **limitación seria de ChatGPT**: cuando le pedimos propuestas concretas de acción, nos dio generalidades.

**[Nuestra propuesta - más allá de lo que sugirió la IA]** Tras debatir entre nosotros, creemos que necesitamos: - **Regulación vinculante YA**, no 'marcos éticos' voluntarios - **Transparencia radical**: saber qué datos usan, cómo deciden los algoritmos - **Auditorías independientes** de sistemas de IA críticos - **Educación masiva** en pensamiento crítico sobre IA

**[Cierre reflexivo]** Senén Barro confiesa ser pesimista pero trabaja para equivocarse. Nosotros, después de este ejercicio de usar IA para analizar IA, entendemos su pes-

imismo: **los modelos actuales son brillantes generando texto, pero carecen de la urgencia moral que esto requiere.**

La pregunta final no es qué puede hacer la IA, sino **qué debemos hacer nosotros con ella**. Y esa respuesta, ChatGPT no puede dárnosla. Gracias.”

---

## **PREPARACIÓN PARA EL DEBATE (5 minutos)**

**Posibles preguntas y respuestas preparadas:**

**Pregunta: “¿Qué alucinaciones específicas detectaron en ChatGPT?”**

- **Respuesta (Persona 1):** “No fueron alucinaciones de datos falsos, sino lo que llamamos **alucinación de control**: el modelo sugería que con ‘gobernanza adecuada’ y ‘marcos éticos’ todo se resolvería, pero no explicaba cómo implementarlos contra intereses económicos masivos. Nos dio **soluciones sin mecanismos de ejecución.**”

**Pregunta: “¿En qué coincidieron con ChatGPT?”**

- **Respuesta (Persona 2):** “En la identificación de riesgos técnicos básicos: deep-fakes, sesgos algorítmicos, pérdida de empleos. Pero discreparamos en la **urgencia y magnitud**. El modelo los presentaba como problemas ‘gestionables’; nosotros creemos que son crisis estructurales.”

**Pregunta: “¿ChatGPT les ayudó o les limitó?”**

- **Respuesta (Persona 3):** “Ambas cosas. Nos ayudó a **estructurar ideas y explorar ángulos** que no habíamos considerado. Pero también nos **empujaba hacia respuestas equilibradas y tibias**. Las ideas más radicales o críticas salieron de nuestros debates, no del modelo. Es como un asistente conservador.”

**Pregunta: “¿Qué le preguntaron exactamente a ChatGPT?”**

- **Respuesta (Persona 1):** “Preguntas como: ‘¿Cuáles son los mayores riesgos de la IA?’, ‘¿Cómo debería ser una IA ética?’, ‘¿Puede la IA acabar con empleos sin crear alternativas?’, ‘¿Quién controla realmente la IA?’. Cada respuesta generaba nuevas preguntas críticas entre nosotros.”

**Pregunta: “¿Detectaron sesgos en las respuestas del modelo?”**

- **Respuesta (Persona 2):** “Sí, un **sesgo pro-tecnológico**: tendía a asumir que más IA es mejor si se hace ‘correctamente’. Nunca sugirió que quizás **no deberíamos desarrollar ciertas aplicaciones**, solo que deberíamos hacerlo ‘éticamente’. También detectamos **sesgo de clase**: las soluciones asumían acceso a educación y recursos.”

### **Pregunta: “¿Confían en que la IA pueda autorregularse?”**

- **Respuesta (Persona 3):** “No. Y esto fue una discrepancia clara con ChatGPT. El modelo sugería mecanismos de ‘alineación’ y ‘entrenamiento ético’. Nosotros creemos que es **ingenuo** esperar que sistemas diseñados para maximizar engagement o beneficios se autolimiten. Necesitamos **regulación externa y dura.**”

### **Pregunta: “¿Qué opinan de que Senén Barro sea pesimista?”**

- **Respuesta (Persona 1):** “Es coherente. Él ve de primera mano cómo funciona la investigación en IA: presiones por publicar, financiación privada con intereses, carrera tecnológica entre países. ChatGPT nos dio optimismo; Barro nos da **realismo.** Preferimos el realismo.”
- 

### **DIPOSITIVA ÚNICA (Apoyo Visual)**

**Título:** LA IA EN EL AÑO 6000: Entre la Distopía y la Esperanza

**Sección Superior: NUESTRO PROCESO:** - □ Debate con ChatGPT - △ Alucinación detectada: “ilusión de control” - □ Análisis crítico grupal - □ Discrepancias: urgencia, soluciones reales

**Columna izquierda - RIESGOS:** - □ Armas autónomas (YA existen) - □ Deepfakes indetectables - ✖ Concentración de poder - □ Externalización de pensamiento

**Columna derecha - FUTURO DESEABLE:** - □ Útil y ética (¿quién decide?) - □ Ubicua pero silente (contradicción) - □ Potenciadora, no sustitutiva - □ Democrática (¿cómo?)

**Centro:** Cita de Senén Barro: “*Soy pesimista sobre las posibilidades de este futuro, pero trabajo cada día para equivocarme.*”

**Pie:** “ChatGPT generó estructura. Nosotros generamos crítica.”

---

### **CONSEJOS ADICIONALES PARA LA PRESENTACIÓN:**

#### **Meta-reflexión constante:**

- Mencionad explícitamente cuándo una idea vino de ChatGPT vs. de vuestro debate
- Señalad contradicciones o limitaciones que detectasteis
- No tengáis miedo de decir “aquí ChatGPT se equivocó” o “esto no nos convenció”

#### **Ejemplos concretos de interacción:**

- Si es posible, llevad **capturas de pantalla** de alguna respuesta particularmente problemática o interesante de ChatGPT
- Mostrad cómo **reformulasteis preguntas** para obtener mejores respuestas

### **Timing ajustado:**

- Persona 1: 1:40 (introducción + metodología)
- Persona 2: 1:40 (riesgos + crítica)
- Persona 3: 1:40 (futuro + conclusión)
- **Total: 5:00 exactos**

### **Lenguaje corporal para el debate:**

- Cuando mencionéis “debatimos entre nosotros”, miraos brevemente
- Mostrad **dinamismo**: que se note que realmente discutisteis estos temas

¡Esto cumple perfectamente con el requisito de análisis crítico del modelo! ☐