

# HACKATHON GUARANÍ 2026

Datos Sintéticos para la IA  
07 · 08 · 09 de agosto de 2026

## GUÍA DEL PARTICIPANTE

Campus UCAP · Hernandarias, Alto Paraná, Paraguay

### ¡Bienvenido/a al Hackathon Guaraní 2026!

Esta guía es tu compañera durante todo el evento. Aquí encontrarás explicaciones claras sobre qué vas a hacer, cómo participar y qué se espera de vos y tu equipo. Léela con calma y sin apuro.

#### 🌟 ¿Qué es este Hackathon?

Durante 3 días (7, 8 y 9 de agosto de 2026), equipos interdisciplinarios trabajarán juntos para crear datasets sintéticos en lengua guaraní. Estos datos serán usados para entrenar modelos de Inteligencia Artificial capaces de entender y hablar guaraní.

No es un concurso de programación pura. Necesitamos personas que hablen guaraní, que trabajen con datos y que sepan de tecnología. Todos los perfiles son igual de importantes.

#### ¿Qué son los Datos Sintéticos?

Antes de arrancar, es fundamental entender con qué vamos a trabajar. Este es el concepto central de todo el Hackathon.

#### La explicación simple

Imaginate que querés enseñarle a una computadora a hablar guaraní. Para eso, necesitás darle miles y miles de ejemplos de frases, conversaciones, preguntas y respuestas en guaraní.

El problema es que no existen suficientes textos digitales en guaraní. Los libros están impresos, las conversaciones se dan de forma oral, y lo que hay en internet es muy poco.

Entonces, ¿qué hacemos? Creamos datos sintéticos: textos generados artificialmente que imitan cómo habla una persona real. No son textos copiados de un libro, sino textos nuevos, creados siguiendo las reglas y el estilo del guaraní real.

#### Una analogía para entenderlo mejor

##### 🗣️ Ejemplo cotidiano






Pensalo como un actor que ensaya para una obra de teatro.

El actor no vivió la historia real, pero aprende a actuar como si la hubiera vivido. Practica los gestos, el tono, las palabras... hasta que parece auténtico.

Los datos sintéticos hacen algo similar: son "ensayos" de cómo se hablaría en guaraní en distintas situaciones — en un hospital, en una oficina, entre amigos — creados de forma artificial pero con tanta calidad que la IA no puede distinguirlos de textos reales.

## ¿Por qué son tan importantes?

Los datos sintéticos son la solución más rápida y eficiente para que el guaraní entre al mundo digital. Veamos por qué:

 <b>Velocidad</b>	No hay que esperar décadas para digitalizar libros o transcribir conversaciones. Se pueden generar miles de ejemplos en horas.
 <b>Privacidad</b>	No usamos datos reales de personas. Todo es generado, así que no hay riesgo de exponer información privada.
 <b>Variedad</b>	Se pueden crear ejemplos en distintas situaciones: salud, educación, gobierno, vida cotidiana. Cuanta más variedad, más inteligente aprende la IA.
 <b>Soberanía</b>	Si los paraguayos generan los datos, la IA aprende el guaraní tal como lo hablamos: con nuestros modismos, nuestra gramática y nuestro contexto cultural.
 <b>Equidad</b>	Permite que hablantes de guaraní accedan a herramientas de IA en su propio idioma, sin quedar excluidos del avance tecnológico.

## ¿Qué pasa si la IA no tiene datos de calidad?

### ⚠ El riesgo de los datos de mala calidad

Si la IA aprende con textos mal escritos, con errores gramaticales o con traducciones literales del español... también hablará con errores.

*Por eso, en este Hackathon el rol del Linguist Hero es tan valioso como el del programador: si el guaraní no es correcto, el dato no sirve.*

# El Proyecto GuaranIA

"GuaranIA: Integrando el Idioma Guaraní en el Ámbito Digital para la Inclusión de Poblaciones Rurales y Vulnerables", Implementado por el CIDIT y apoyado por el BID Lab (Laboratorio del Banco Interamericano de Desarrollo).

## ¿Cuál es la finalidad del proyecto?

GuaranIA nace de una necesidad concreta: miles de paraguayos que viven en zonas rurales o en situación de vulnerabilidad se comunican principalmente en guaraní. Sin embargo, casi todas las herramientas digitales — traductores, asistentes de voz, chatbots de salud, sistemas de atención ciudadana — están diseñadas para funcionar en español o inglés.

Eso genera una barrera real: si no hablás bien español, no podés acceder a esos servicios digitales. GuaranIA quiere romper esa barrera.

## ¿Por qué es urgente?

Para que una IA aprenda a hablar guaraní, necesita miles de ejemplos de textos correctamente escritos. Ese es el problema: hoy no existen suficientes datos digitales en guaraní. Sin datos, no hay IA. Sin IA, los hablantes de guaraní quedan excluidos del avance tecnológico.

<b>El problema</b>	El guaraní tiene millones de hablantes, pero es considerado una lengua de "bajos recursos" en IA por falta de datos digitales.
<b>La solución</b>	Crear datos sintéticos de alta calidad que permitan entrenar modelos de IA capaces de entender y hablar guaraní.
<b>El impacto</b>	Poblaciones rurales y vulnerables podrán acceder a servicios digitales en su idioma nativo por primera vez.
<b>Tu aporte</b>	Los datasets que generes en este Hackathon serán el insumo directo para entrenar esas herramientas.

## ¿Por qué el guaraní es un desafío tecnológico?

Aunque la mayoría de los paraguayos habla guaraní, la IA lo trata como una lengua de "bajos recursos". Eso significa que los modelos de IA saben muy poco de guaraní comparado con el inglés o el español.

### Dato clave

El guaraní tiene millones de hablantes, pero en internet casi no hay textos digitalizados en guaraní. Eso hace que los sistemas de IA fallen o den respuestas incorrectas cuando alguien intenta usarlos en este idioma.

*Tu trabajo en este Hackathon es parte directa de la solución a este problema real.*

## ¿Quién puede participar?

Este Hackathon es abierto e inclusivo. No necesitás ser programador para participar. Lo que se busca es diversidad de perfiles que trabajen en equipo.

### Perfiles habilitados

---

- Estudiantes de carreras tecnológicas de la UCAP u otras universidades de Alto Paraná.
- Estudiantes o graduados en lengua guaraní de cualquier institución reconocida.
- Alumnos del último año de secundaria técnica (BATIN).
- Egresados o participantes del proyecto Desarrollo Joven.

#### **Importante: menores de edad**

Si tenés menos de 18 años, necesitás presentar una autorización escrita y firmada por tu padre, madre o tutor legal al momento de la acreditación. Sin ese documento, no podrás ingresar al evento.

## Los Equipos y sus Roles

Los equipos deben tener entre 3 y 5 personas. La clave es que cada equipo sea interdisciplinario: se necesitan diferentes habilidades trabajando juntas. También existe una Categoría Colegio para instituciones educativas de nivel secundario, y una modalidad excepcional de inscripción individual. Ver detalle más adelante.

### Los tres roles esenciales

---

#### **Data Wrangler — El/La Organizador/a**

Perfil: Estudiante del último año de bachillerato técnico (BATIN) con interés en tecnología.

¿Qué hace?

- Organiza y estructura los datos según el formato oficial del evento.
- Elimina duplicados, normaliza caracteres y verifica formatos.
- Colabora con el Lead Dev en la preparación de los archivos finales.

Habilidades útiles: manejo básico de Excel/CSV, Google Sheets, atención al detalle.

#### **Linguist Hero — El/La Guardián/a del Guaraní**

Perfil: Estudiante, graduado/a o docente de lengua guaraní con dominio avanzado del idioma.

¿Qué hace?

- Garantiza que los textos generados sean gramaticalmente correctos en guaraní.
- Verifica que el lenguaje sea natural y fluido, no una traducción literal del español.
- Incorpora modismos, expresiones regionales y variaciones dialectales.
- Valida los datos generados automáticamente antes de la entrega final.

Habilidades útiles: dominio escrito y oral del guaraní, gramática formal, sensibilidad cultural.

## **Lead Dev — El/La Desarrollador/a**

Perfil: Estudiante avanzado/a o graduado/a de Informática, Sistemas o carrera afín.

¿Qué hace?

- Diseña y ejecuta los scripts para generar los datos sintéticos.
- Implementa pipelines para transformar, filtrar y exportar los datasets.
- Coordina con el Data Wrangler y el Linguist Hero.
- Documenta el código para la presentación ante el jurado.

Habilidades útiles: Python (Pandas, HuggingFace), manejo de JSON/CSV, modelos de lenguaje.

## **Límite de equipos**




Solo habrá 5 equipos en total. Puede haber más de una persona con el mismo perfil dentro del equipo, siempre que haya al menos un representante de cada rol.

La organización puede sugerir recomposiciones de equipo antes del cierre de inscripciones para garantizar el perfil interdisciplinario.

## **Categoría Colegio**

¡Este Hackathon tiene una categoría especial para colegios! Si sos estudiante o docente de una institución educativa de nivel secundario, podés armar un equipo con compañeros de tu mismo colegio y competir por el reconocimiento institucional.

### **Premio Categoría Colegio — Primer Puesto:**

-  Trofeo institucional para el colegio ganador.
-  Reconocimiento oficial para la institución ganadora.
-  Premio económico de Gs. 2.000.000 distribuido en partes iguales entre los participantes del equipo ganador.

Los equipos de la Categoría Colegio deben cumplir con los mismos requisitos de conformación, roles y participación que el resto de los equipos.




## **Inscripción Individual — Modalidad Excepcional**

Si querés participar pero no tenés un equipo formado, existe una opción excepcional: podés inscribirte de forma individual postulando a uno de los tres roles (Data Wrangler, Lead Dev o Linguist Hero). La organización intentará conformar un equipo completo agrupando a los participantes individuales que hayan solicitado roles complementarios.

### **Importantísimo — Leé esto antes de inscribirte de forma individual**

Esta modalidad es **excepcional y no garantiza participación**. Si la organización no logra conformar un equipo completo con los tres roles necesarios, los participantes inscritos de forma individual **no podrán participar del Hackathon**. La organización notificará a cada participante individual con anticipación si el equipo pudo conformarse o no.

## Inscripción: Cómo Anotarse

 <b>Apertura</b>	23 de abril de 2026
 <b>Cierre</b>	25 de julio de 2026 (improrrogable)
 <b>Modalidad</b>	Formulario oficial en los canales digitales de UCAP y CIDIT <a href="https://inscripciones.ciditpy.org/">https://inscripciones.ciditpy.org/</a>

### ¿Qué necesitas para inscribirte?

- Nombre del equipo (único e irrepitable).
- Datos de todos los integrantes: nombre completo, CI, institución, carrera/año, email y contacto.
- Rol asignado a cada integrante.
- Designar un líder o representante del equipo.
- Declaración de aceptación de las bases (firmada por cada integrante o su tutor si es menor de edad).

#### ¡Atención!

Las inscripciones cierran improrrogablemente el 25 de julio de 2026 y estas se realizan en línea.

No se aceptarán postulaciones fuera de plazo ni cambios en la composición del equipo una vez confirmada la inscripción, salvo casos de fuerza mayor debidamente justificados.

#### Confirmación de inscripción




La organización notificará la aceptación o rechazo dentro de los 3 días hábiles posteriores al cierre del plazo.

Los equipos aceptados recibirán por email las credenciales para la Onboarding Week y los materiales de preparación.

## Onboarding Week (Pre-Hackathon)

Del 3 al 5 de agosto de 2026, antes del evento principal, participarás de una semana de preparación virtual. Es fundamental para que todos los equipos arranquen en igualdad de condiciones.

### ¿Qué se hace durante la Onboarding Week?

 <b>Webinars técnicos</b>	Sesiones sobre herramientas, modelos de lenguaje y técnicas de generación de datos sintéticos.
 <b>Presentación de retos</b>	Explicación detallada del desafío central para que los equipos empiecen a idear soluciones.
 <b>Acceso a materiales</b>	Entrega de datasets base, guías de estilo lingüístico y documentación técnica.

## ¿Quién debe asistir?

---

### **Obligatoriedad de asistencia — MUY IMPORTANTE**

WEBINARS TÉCNICOS: Debe asistir obligatoriamente el líder del equipo (o un miembro nominado con anticipación). La ausencia injustificada es causa de descalificación.


SESIÓN DE PRESENTACIÓN DE RETOS: Deben asistir TODOS los integrantes del equipo. Si falta alguno sin justificación aceptada, el equipo queda inhabilitado para participar.

Las justificaciones deben enviarse por email a la coordinación con al menos 24 horas de anticipación, con documentación de respaldo. La organización resuelve cada caso de forma definitiva.

## Programa Día a Día

### Día 1 — Viernes 7 de agosto de 2026

---

Horario	Actividad
08:00 – 09:00	Acreditación y bienvenida. Entrega de kits a los equipos.
09:00 – 09:30	Apertura oficial: palabras de autoridades de la UCAP y CIDIT.
09:30 – 10:00	Presentación del proyecto GuaranIA y el equipo técnico de UDELAR.
10:00 – 10:30	Explicación de reglas, criterios de evaluación y entrega de datasets base.
10:30 – 11:00	Café de networking — integración de equipos.
11:00 – 13:00	 Inicio formal del desarrollo. Equipos trabajan en sus soluciones.
13:00 – 14:00	Almuerzo.
14:00 – 18:00	Desarrollo intensivo. Rondas de mentoría individual por equipo.
18:00 – 18:30	Check-in vespertino: cada equipo comparte un avance breve (2 min).
18:30 – 20:00	Trabajo libre y resolución de dudas técnicas.



### Día 2 — Sábado 8 de agosto de 2026


---

Horario	Actividad
08:00 – 09:00	Desayuno y reanudación de actividades.
09:00 – 10:30	Charla técnica: Buenas prácticas en generación de datos sintéticos para PLN.
10:30 – 13:00	Desarrollo. Segunda ronda de mentoría individual por equipo.
13:00 – 14:00	Almuerzo.
14:30 – 18:00	Desarrollo intensivo. Ajuste y validación lingüística de los datasets.
18:00 – 18:30	Check-in vespertino: presentación de avances (3 min por equipo).
18:30 – 20:00	Trabajo libre. Preparación de la presentación final.

### Día 3 — Domingo 9 de agosto de 2026

---

Horario	Actividad
08:00 – 09:00	Desayuno y última sesión de trabajo.
09:00 – 10:30	Cierre del desarrollo. Preparación y ensayo de presentaciones finales.
10:30 – 10:45	 Entrega oficial de proyectos y datasets a la organización.
10:45 – 11:00	Receso.
11:00 – 13:00	 Demo Day: presentaciones finales ante el jurado (10 min + 5 min de preguntas).

<b>13:00 – 14:00</b>	Almuerzo. Deliberación del jurado.
<b>14:00 – 14:45</b>	 Anuncio de ganadores y entrega de premios.
<b>14:45 – 15:00</b>	Cierre oficial y palabras finales.
<b>15:00</b>	Clausura y foto grupal.

## ¿Qué tienen que Entregar?

Al finalizar el Hackathon (el domingo 9 de agosto a las 10:30), cada equipo debe entregar los siguientes entregables a la organización:

### Entregables obligatorios

1. Dataset sintético en guaraní en el formato indicado en la guía de estilo oficial.
2. Informe técnico (máximo 5 páginas) describiendo:
  - La metodología utilizada.
  - Las herramientas empleadas.
  - Los resultados obtenidos.
3. Código fuente o scripts utilizados, documentados y bajo licencia abierta (MIT o CC BY 4.0).
4. Presentación digital (máximo 10 diapositivas) para el Demo Day ante el jurado.

## El Demo Day — La presentación final

Durante el Demo Day, cada equipo tendrá:

- 10 minutos para presentar la solución.
- 5 minutos para responder preguntas del jurado.

La presentación debe incluir una demostración del dataset y explicación de la metodología.

### Asistencia obligatoria

Todos los integrantes del equipo deben estar presentes durante la presentación.

## ¿Cómo se Evalúa?

El jurado evaluará cada proyecto sobre 100 puntos, considerando estos cinco criterios:





Criterio de Evaluación	Peso	Puntaje
Calidad y fidelidad lingüística del dataset generado	30%	30 pts
Volumen y diversidad de los datos producidos	20%	20 pts
Innovación en la metodología o técnica de generación	20%	20 pts
Potencial de uso real (académico o de mercado)	15%	15 pts
Presentación, documentación y trabajo en equipo	15%	15 pts
<b>TOTAL</b>	<b>100%</b>	<b>100 pts</b>

### Tip clave: la calidad lingüística vale el 30%

- El criterio más importante es la calidad del guaraní en el dataset.
- Eso significa que el trabajo del Linguist Hero es decisivo para ganar.

- Un dataset grande pero con errores lingüísticos vale mucho menos que uno más pequeño pero perfecto.
- En caso de empate entre equipos, el jurado dará preferencia al que tenga mayor calidad lingüística.

## Premios

 <b>1.er Puesto</b>	Gs. 2.500.000 + Certificado de excelencia + Trofeo + Mención en publicaciones del programa GuaranIA (BID Lab)
 <b>2.do Puesto</b>	Gs. 2.000.000 + Certificado de reconocimiento + Trofeo
 <b>1.er Puesto Categoría Colegio</b>	Troféo institucional para el colegio + Reconocimiento oficial + Gs. 2.000.000 distribuido entre los participantes del equipo
 <b>Todos los participantes</b>	Certificado oficial de participación en el Hackathon Guaraní 2026

### ¿Cómo se reparte el premio?

El monto en efectivo se divide en partes iguales entre todos los integrantes del equipo ganador.

El pago se realiza mediante transferencia bancaria, a nombre del líder del equipo, dentro de los 15 días hábiles posteriores al evento.

Para cobrar, todos los integrantes deben presentar su documento de identidad vigente y haber firmado la cesión de derechos sobre el dataset.

## Propiedad Intelectual

Este es un punto importante que hay que entender antes de participar.

### ¿Qué pasa con lo que creen?

**DATASETS:** Los datasets generados durante el Hackathon serán cedidos al programa GuaranIA para entrenar modelos de IA. Esto es la razón de ser del evento.

**AUTORÍA:** Los equipos siempre serán reconocidos como autores en publicaciones y usos derivados de sus datos.

**CÓDIGO:** El código fuente queda disponible bajo licencia abierta (MIT o CC BY 4.0). Esto significa que cualquiera puede usarlo, modificarlo y distribuirlo, siempre que mencione a los autores.

## Reglas que Hay que Conocer

### Lo que está permitido

---

- Consultar a los mentores designados por la organización.
- Usar herramientas y datasets provistos por la organización.
- Trabajar en equipo de forma colaborativa dentro del propio grupo.
- Buscar información técnica en internet (documentación, tutoriales públicos).

### Lo que NO está permitido

---

- Colaborar técnicamente con otros equipos durante la competencia.
- Incorporar datasets externos no provistos ni autorizados por la organización.
- Copiar o plagiar código, datasets o metodologías sin declararlo.
- Intentar influir indebidamente en las decisiones del jurado.
- Falsificar datos personales en la inscripción.
- Conductas irrespetuosas hacia otros participantes, mentores, jurado u organizadores.

#### Causales de descalificación automática

- Ausencia injustificada del líder a los webinars de la Onboarding Week.
- Ausencia injustificada de cualquier integrante a la sesión de presentación de retos.
- Plagio o uso de datos no autorizados.
- Conducta irrespetuosa.
- Falsificación de datos en la inscripción.

En caso de descalificación, el equipo no tendrá derecho a reclamo sobre premios ni reconocimientos.

## **? Preguntas Frecuentes**

### **¿Necesito saber programar para participar?**

No. Si sos estudiante de guaraní o alumno del BATIN, tu rol es igualmente esencial. Lo que se necesita es que el equipo en conjunto tenga todas las habilidades.

### **¿Puedo inscribirme sin equipo completo?**

Sí, existe una modalidad excepcional de inscripción individual. Podés registrarte postulando a uno de los tres roles (Data Wrangler, Lead Dev o Linguist Hero) y la organización intentará conformar un equipo con los participantes individuales de roles complementarios. Importante: si no se logra conformar un equipo completo con los tres roles, no podrás participar. La organización te avisará con anticipación. Si querés más seguridad, lo ideal es inscribirte ya con un equipo formado.

### **¿Qué pasa si no puedo asistir a alguna actividad de la Onboarding Week?**

Debés justificar la ausencia por email a la coordinación técnica con al menos 24 horas de anticipación y con documentación de respaldo. La organización decidirá en cada caso.

### **¿Necesito llevar mi propia computadora?**

Se recomienda que el Lead Dev y el Data Wrangler lleven sus propias laptops. La organización brindará información sobre equipamiento disponible antes del evento.



### **¿El Hackathon es presencial o virtual?**

La Onboarding Week (3–5 de agosto) es virtual. El Hackathon principal (7–9 de agosto) es presencial en el Campus UCAP, Hernandarias, Alto Paraná.

### **¿El código que hago se puede usar comercialmente?**

El código queda bajo licencia abierta (MIT o CC BY 4.0), lo que permite su uso, modificación y distribución, incluso comercialmente, siempre que se mencione a los autores originales.

## **Contacto e Información**

 <b>Organización principal</b>	Universidad Católica de Alto Paraná (UCAP) — Facultad de Ciencias y Tecnología · Campus Hernandarias, Alto Paraná, Paraguay - CIDIT — Centro de Ingeniería para la Investigación, Desarrollo e Innovación Tecnológica — Proyecto GuaranIA (BID Lab)
 <b>Coordinación académica</b>	Luis Chiruzzo y Agustín Millán — Universidad de la República (UDELAR), Uruguay

### **17 Fechas clave de un vistazo**

- 23 abril – 25 julio 2026 → Período de inscripciones
- 3 – 5 agosto 2026 → Onboarding Week (virtual, obligatoria)
- 7 – 9 agosto 2026 → Hackathon principal (presencial, Hernandarias)
- 9 agosto 2026, 10:30 hs → Entrega final de proyectos
- 9 agosto 2026, 14:00 hs → Anuncio de ganadores

***¡Ñañemoñe'ë guaraníme! · ¡Hablemos en guaraní con la IA!***

UCAP · CIDIT · GuaranIA · UDELAR · BID Lab

*Hernandarias, Alto Paraná, Paraguay*