

PILLAR

INSIGHTS

AI image generators struggle with non-English text — and reinforce linguistic inequality.

A first-hand account of asking AI image models to write Spanish, French, and Portuguese inside their outputs — and watching the language fall apart in ways that map directly to phonetic, regional, and dataset bias.

PILLAR MEDIA & ENTERTAINMENT · PILLARME.COM/INSIGHTS

01 — The illusion of competence.


One of the main challenges I encountered while working with AI systems was the issue of language barriers and linguistic inaccuracies, especially in AI image generation tools. While these technologies are often presented as highly advanced and multilingual, my experience showed that their handling of written language inside generated images remains surprisingly unreliable. This limitation became particularly visible when working with prompts in languages other than English, or when asking the AI to generate precise text within visual content.

At first glance, image generation models appear extremely sophisticated. They can create realistic photographs, artistic illustrations, advertisements, posters, and complex visual compositions in seconds. However, when these same systems are asked to include written language in the image, many inconsistencies begin to appear. Instead of producing coherent words or grammatically correct sentences, the AI often generates distorted typography, invented words, mixed languages, or meaningless combinations of letters that resemble language without actually being understandable.

One of the most interesting aspects of this issue is that the AI frequently gives the *illusion* of linguistic competence. The generated text often looks visually convincing from a distance because it imitates the structure of real writing. The spacing, typography, and visual rhythm resemble actual language, but upon closer inspection, the words are often incorrect or completely fictional. This suggests that the model may not truly process written language inside images as language itself, but rather as a visual pattern or aesthetic texture.

IA en las escuelas latinas





An AI-generated image that looks competent at thumbnail size. Up close, the words dissolve into shapes that imitate writing without being writing.

02 — The “educasion” case study.

This image seemed particularly interesting to me because the spelling mistakes generated by the AI closely resemble common orthographic errors found in parts of Latin America due to pronunciation differences with European Spanish. For example, the AI wrote “**educasion**” instead of “**educación**,” reflecting the way the word is often pronounced in many Latin American accents, where the “c” and “s” sounds are pronounced similarly.

In contrast, in Spain, the pronunciation more clearly distinguishes the “ción” sound. What makes this especially fascinating is that the AI appears to reproduce not only linguistic patterns, but also **phonetic tendencies associated with specific regional varieties of Spanish**. The model has internalized what LATAM Spanish sounds like without internalizing the standardized orthography that goes with it.



“Educasion” in the wild. The misspelling mirrors how the word is pronounced in much of Latin America — evidence that the model absorbed phonetic patterns, not just text.

03 – Multilingual contexts make it worse.

This problem becomes even more pronounced in multilingual contexts. During my experimentation, I noticed that English prompts generally produced more coherent results than prompts written in French, Spanish, or Portuguese. Even when the AI understood the general meaning of the request, the generated text frequently contained spelling mistakes, missing accents, incorrect grammar, or mixtures between several languages at once. In some cases, the AI invented entirely new words that resembled fragments of existing languages but had no actual meaning.

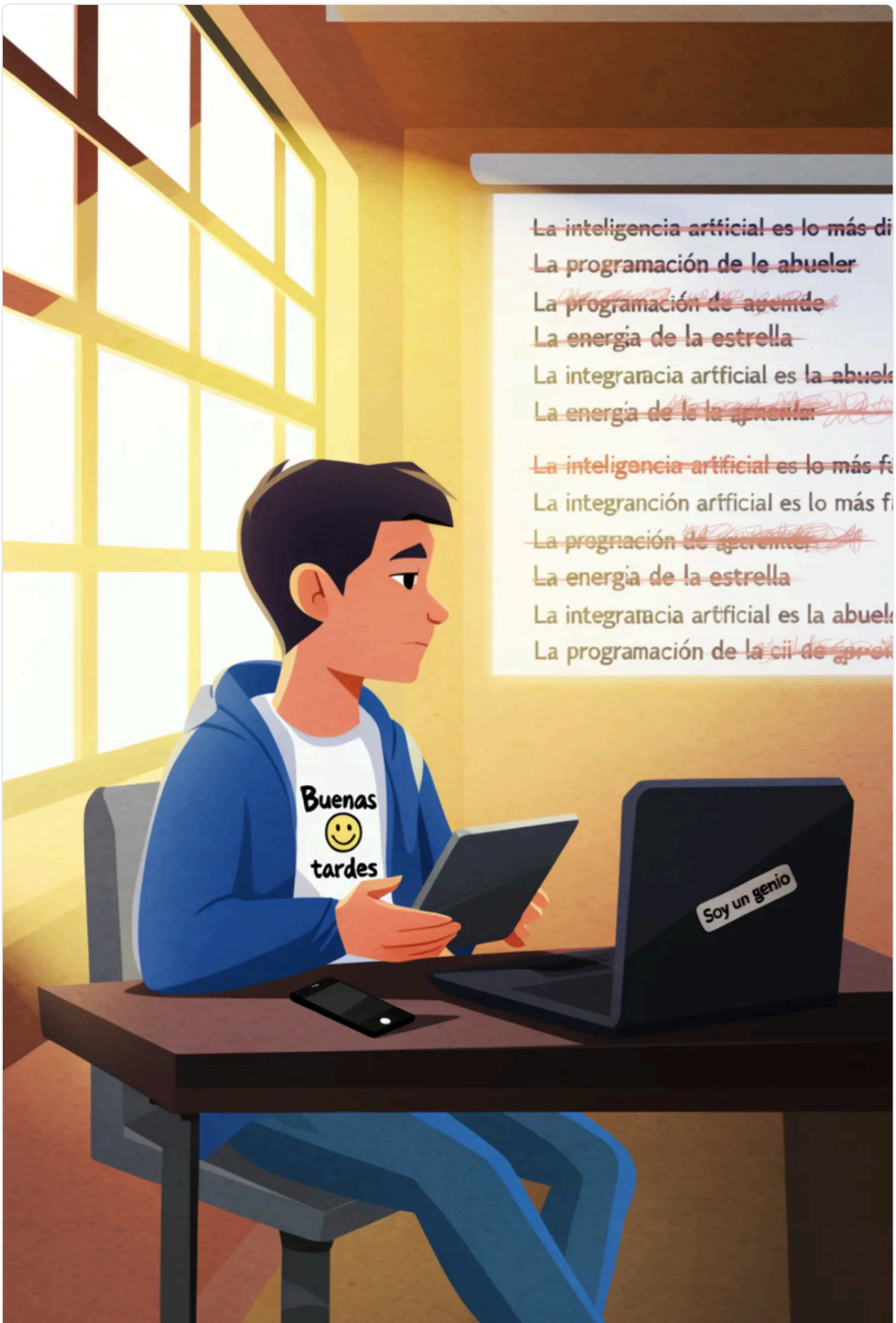


A multilingual prompt produces invented vocabulary and fragments that resemble French and Spanish without being either. The model gestures at language; it does not write it.

04 – The AI “corrects” by crossing out.

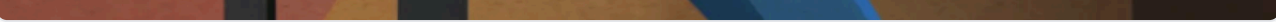
Another example that I found particularly revealing was an image containing completely nonsensical text. After noticing the errors, I asked the AI to correct the writing while keeping the same visual style and structure. However, instead of producing a coherent version, the system generated a second image that *still* contained meaningless text.

Even more interestingly, rather than actually correcting the mistakes, the AI simply **crossed out some of the incorrect words**, as if visually acknowledging the errors without being capable of properly fixing them. This gave the impression that the model recognized something was wrong at a superficial level, while still lacking a true understanding of the linguistic content itself.



~~La inteligencia artificial es lo más di~~
~~La programación de le abueler~~
~~La programación de aprende~~
~~La energía de la estrella~~
~~La integraracia artificial es la abueler~~
~~La energía de lo la aprenenar~~

~~La inteligencia artificial es lo más fe~~
~~La integranción artificial es lo más fi~~
~~La prognación de aprenenar~~
~~La energía de la estrella~~
~~La integraracia artificial es la abueler~~
~~La programación de la cil de apren~~



Asked to fix the text, the model drew strike-throughs over its own mistakes. It recognized something was wrong; it could not say what.

05 – Spanish posters, invented vocabulary.

When requesting the creation of a Spanish poster, the AI generated sentences that visually resembled Spanish but included impossible grammatical structures and invented vocabulary. Additionally, it sometimes resulted in hybrid text mixing Spanish and English elements together. This inconsistency revealed an important limitation: although AI systems are marketed as multilingual tools, their performance remains strongly influenced by **the dominance of English-language data in training datasets**.



PROYECTO
IA PARA ESCUELAS LATAM

Inteligencia Artificial al servicio de la educación.

Herramientas de IA para directivos y docentes

Nuestro proyecto impulsa la transformación educativa en Latinoamérica, brindando herramientas de IA seguras, prácticas y



Perfeccionar imágenes



Editar imagen completa

Toca Editar y describe los cambios en el chat.



Selecciona un área de la imagen para editarla

Toca Seleccionar, destaca el área y describe los cambios en el chat.

Comenzar

Obtener más información

A "Spanish poster" from an image model: real-looking layout, broken grammar, half-Spanish half-English vocabulary. The visual frame is correct. The language is not.

06 – Unpredictability.

Another major issue was the unpredictability of the results. The exact same prompt could produce drastically different outputs from one generation to another. In some attempts, the text was correct or almost correct, while in others it became completely unreadable. This inconsistency made the tools difficult to rely on for professional or academic use. Even after refining prompts multiple times, adding specifications, or simplifying instructions, the AI continued to struggle with Spanish textual precision inside images.

The model knows what writing looks like. It does not know what writing means.

07 — Practical consequences for users.

This limitation has important practical consequences. A poster, infographic, advertisement, or presentation containing distorted or incorrect language may appear unprofessional or even misleading. As a result, users often need to manually edit the generated images afterward using external software. The AI still requires human correction.

However, with the Pro version of ChatGPT, you unlock many more powerful options to edit photos and images with ease. From improving image quality and correcting errors to generating professional designs and making advanced creative adjustments, the process becomes faster, simpler, and much more flexible — though the underlying language problem inside the image itself remains.



PROYECTO
IA PARA ESCUELAS LATAM

Inteligencia Artificial al servicio de la educación.

Herramientas de IA para directivos y docentes

Nuestro proyecto impulsa la transformación educativa en Latinoamérica, brindando herramientas de IA seguras, prácticas y adaptadas a la realidad de cada escuela.



Ahorra tiempo en tareas administrativas



Toma decisiones basadas en datos



Mejora los aprendizajes y la experiencia educativa



IA segura, ética y alineada a valores educativos



MEJORES DECISIONES
MÁS TIEMPO PARA LO QUE REALMENTE IMPORTA:
NUESTROS ESTUDIANTES



Escuelas más eficientes, equipos más empoderados, estudiantes **con más oportunidades.**



Realistic-looking infographic typography that falls apart at full resolution. For professional use the output still needs human correction in external software.

o8 – The broader question: accessibility and representation.

The issue also raises broader questions about accessibility and representation. Languages with smaller digital footprints or less representation in AI training datasets may be more vulnerable to poor-quality outputs. This creates an imbalance where English-speaking users benefit from more accurate and optimized systems, while multilingual users experience higher rates of errors and limitations. In this sense, AI systems may unintentionally reinforce **existing linguistic inequalities in digital spaces**.

Another interesting observation is the contrast between textual AI systems and image-generation AI systems. Text-based models are often capable of producing highly coherent multilingual writing, while image generators struggle to reproduce even simple sentences accurately. This suggests that generating language visually is fundamentally different from generating language textually.

Screenshots of incorrect generations revealed patterns that would have been difficult to explain theoretically. Some images contained signs with almost-correct words, while others produced typography that looked realistic but was entirely unreadable. In several cases, the AI ignored explicit instructions regarding language choice or spelling accuracy. These visual failures demonstrated that despite the impressive appearance of generative AI systems, significant technical limitations remain unresolved.

Cancelar

Seleccionar

Siguiente



PROYECTO
IA PARA ESCUELAS LATAM

Inteligencia Artificial al servicio de la educación.

Herramientas de IA para directivos y docentes

Nuestro proyecto impulsa la transformación educativa en Latinoamérica, brindando herramientas de IA seguras, prácticas y adaptadas a la realidad de cada escuela.



Ahorra tiempo en tareas administrativas



Toma decisiones basadas en datos



Mejora los aprendizajes y la experiencia educativa

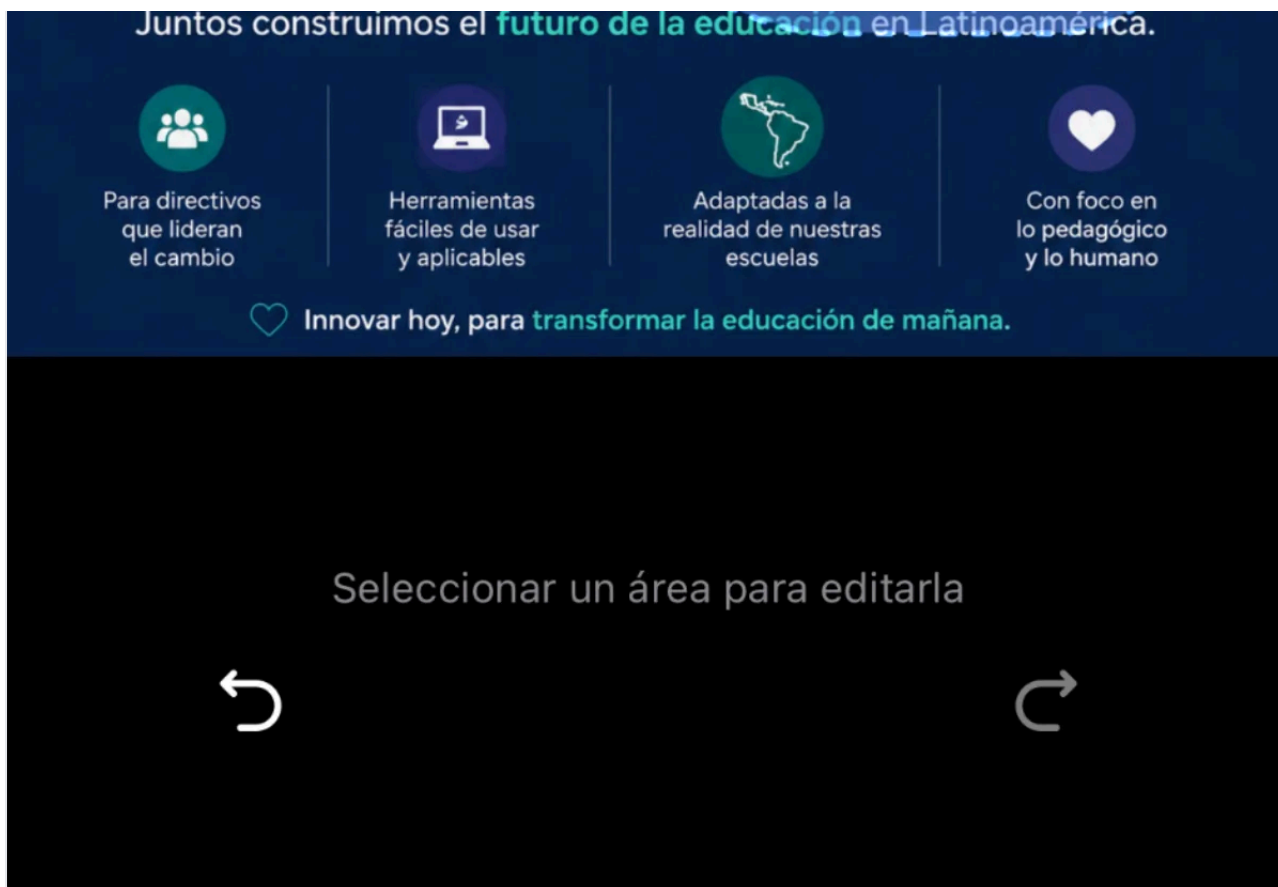


IA segura, ética y alineada a valores educativos



Escuelas más eficientes, equipos más empoderados, estudiantes **con más oportunidades.**

MEJORES DECISIONES
MÁS TIEMPO PARA LO QUE REALMENTE IMPORTA:
NUESTROS ESTUDIANTES



Explicit language and spelling instructions, ignored. The technical limitation is not aesthetic — it is linguistic, and it is unevenly distributed across the world's languages.

Ultimately, this experience changed my perception of AI reliability. While AI is often presented as a universal solution capable of replacing human labor across creative industries, these experiments showed that current systems still struggle with linguistic precision, especially in multilingual and multimodal contexts. This highlights the importance of using AI critically and carefully verifying outputs, particularly when working across different languages.

At the same time, I was genuinely surprised by how impressive many of the generated images and creative results were. The technology is undeniably powerful and evolving rapidly, even if challenges around multilingual accuracy and textual reliability still remain important areas for improvement.

ABOUT THE AUTHOR

Marie Prieto

Marie Prieto is a researcher and writer focused on how AI systems perform — and fail — across the languages of the Spanish- and French-speaking world. This piece is a LATAM Ambassador contribution to Pillar Insights.

[Read more Pillar Insights \(/insights\)](#)

Frequently asked questions.

Why do AI image generators struggle with text inside images?

AI image generators do not appear to process written language inside images as language itself. They treat letters and words as visual patterns or aesthetic textures. From a distance the generated text looks convincing because the spacing, typography, and visual rhythm imitate the structure of real writing. On closer inspection the words are frequently incorrect, garbled, or completely fictional. This is why posters and signage from image models can look credible at thumbnail size and fall apart at full resolution. The model knows what writing looks like, but it does not know what writing means.

Why does AI image generation perform better in English than in Spanish, French, or Portuguese?

English prompts generally produce more coherent text inside generated images than prompts in French, Spanish, or Portuguese. Even when the AI understands the general meaning of a non-English request, the visual output frequently contains spelling mistakes, missing accents, incorrect grammar, or mixtures of several languages at once. The cause is dataset imbalance. AI image models are heavily influenced by the dominance of English-language data in their training sets, so non-English languages inherit higher rates of errors. AI systems marketed as multilingual still privilege English in practice.

What does the “educasion” misspelling reveal about AI language models?

When an AI image generator wrote “educasion” instead of “educación,” it reproduced a common orthographic error found in parts of Latin America, where seseo pronunciation collapses the c and s sounds together. In peninsular Spanish the ción ending is pronounced more distinctly. The misspelling suggests the model is not just learning written Spanish vocabulary but also absorbing phonetic tendencies of specific regional accents. The AI has learned what LATAM Spanish sounds like without learning the standardized orthography that goes with it.

What happens when you ask AI to correct text inside a generated image?

Asking an AI image model to correct broken text inside a previously generated image rarely produces a coherent fix. In one revealing case the model regenerated an image and, instead of repairing the words, simply drew strike-throughs over the incorrect text. The system appeared to recognize that something was wrong at a superficial level but lacked the linguistic understanding needed to actually fix it. This pattern reinforces the conclusion that image generators do not process embedded text as language. They process it as a visual element that can be replaced or crossed out, not as a sequence of characters with meaning.

Why are AI image outputs so unpredictable from one generation to the next?

The same prompt can produce drastically different outputs from one generation to another. In some attempts the embedded text is correct or nearly correct. In others it is unreadable. Even with refined prompts, additional specifications, or simplified instructions, AI image generators continue to struggle with Spanish textual precision inside images. This inconsistency makes the tools difficult to rely on for professional or academic work without manual human correction afterward in external software.

How does this affect accessibility and representation for non-English languages?

The text-generation gap raises broader questions about accessibility and representation. Languages with smaller digital footprints or less representation in AI training datasets are more vulnerable to poor-quality outputs. English-speaking users benefit from more accurate and optimized systems, while multilingual users experience higher rates of errors and limitations. AI systems can therefore unintentionally reinforce existing linguistic inequalities in digital spaces. This is the same dynamic Pillar is built to counter at the infrastructure level: closing the quality gap between English and the languages of the Global South.

If AI image models can't spell "educación," the same dataset gap is shaping which languages get premium digital real estate. Pillar manages 100,000+ premium properties across Spanish, French, Portuguese, and the languages of the Global South.