

Introducing The Language Infrastructure Framework

A reference model for what it takes for a
language to live end-to-end online



Coalition on
Digital Impact

codi.global

What Does It Mean for a Language to Exist on the Internet? A CODI Draft Proposal for Community Discussion

This document is offered as a draft proposal for discussion. It is not a finished framework, a specification, or a standard. Its purpose is to provide a shared starting point for thinking together about what a meaningful linguistic presence on the Internet actually requires.

Conversations about multilingual access often focus on individual components of the Internet, such as domain names, websites, applications, or content. Each of these matters. But users do not experience the Internet as a set of components. They experience it as a continuous interaction that involves identifying themselves, expressing ideas, navigating systems, finding information, completing tasks, and understanding what is happening to them along the way.

Language is present or absent at every point in that experience. A language can be technically supported in one place and still fail the user in another. It can be readable but not writable, visible but not discoverable, or available for browsing but absent at moments of consequence, such as payments, safety warnings, or account recovery. When that happens, users are quietly forced to abandon their language midstream.

This draft proposal offers a way to describe where language must exist across the full Internet experience in order for a person to meaningfully participate online in their own language.

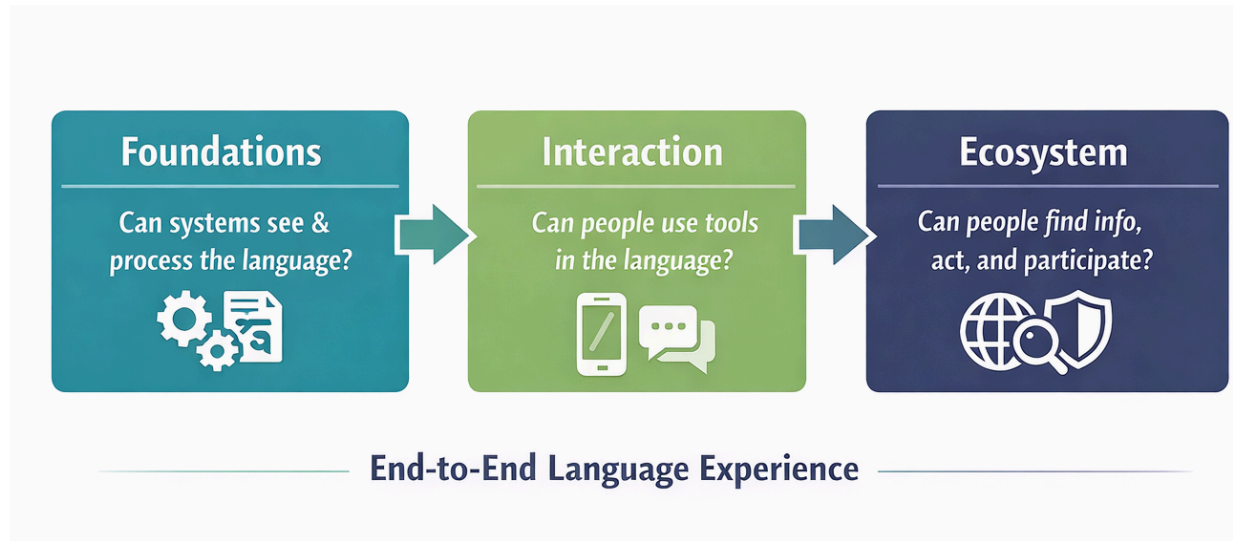
The intent here is to invite critique, refinement, and shared ownership from the community.

The Language Infrastructure Framework

The core assumption of this model is simple: a language is meaningfully present on the Internet only if a person can complete common digital tasks entirely in that language, without being pushed into a fallback they did not choose. In other words, a language must be able to live end-to-end online.

That presence depends on a series of layers. These layers map directly to moments users encounter every day. These layers also fall into three broad phases: the technical foundations that make a language legible and usable by systems, the interfaces where people interact with digital tools, and the broader ecosystem where information, participation, and safety take shape.

Taken together, these layers follow the path of a user online: a language must first be machine-legible, functional for identity and expression, supported in the tools people rely on, understandable at moments of decision, present in content, visible through algorithms, usable for participation, and present where safety and accountability matter.



Technical Foundations

Layer 1: Encoding and Script Legibility

Before anything else can work, a language must be legible to machines. Its script needs to be encodable, renderable, and stable across the environments people actually use. When this layer fails, users encounter broken characters, missing glyphs, or substituted symbols. At that point, the Internet cannot reliably see the language.

This layer is often treated as solved, but in practice, it depends on consistent implementation across platforms, updates, and device classes.

In scope of this layer:

- Unicode encoding and normalization
- Font availability and fallback behavior
- Rendering across major operating systems
- Rendering across modern web browsers
- Display consistency across desktop and mobile devices

An analysis at this level would typically involve reviewing standards bodies, OS and browser implementations, font libraries, and real-world rendering behavior across common devices.

If this layer breaks, every subsequent layer inherits the failure.

Layer 2: Identity and Addressability

Once a language can be rendered, it must be usable for identity. People need to be able to name themselves and be reached in their own language through domain names, email addresses, URLs, usernames, and form fields.

This is where Universal Acceptance sits. If systems reject a language at this layer, users are forced to transliterate, simplify, or abandon their language at the moment they establish an online presence. Inclusion that fails at identity is fragile by definition.

In scope of this layer:

- Internationalized domain names (IDNs)
- Email address internationalization (EAI)
- URLs and link handling
- Usernames and account identifiers
- Form validation and input acceptance

An analysis at this level would typically involve registries, registrars, email providers, DNS operators, platform login systems, and common form-validation behavior across major services.

Layer 3: Input and Expression

Even when a language can be displayed and accepted by systems, users still need to be able to express themselves. This is where many languages quietly fall out of the Internet experience.

Input is not just about keyboards. It is about whether people can reliably type, compose, and correct text using the tools available to them across devices and everyday contexts. A language that can be read but not written is functionally muted. It exists online, but only as something spoken to users, not by them.

In scope of this layer:

- Physical keyboard layouts
- Virtual and mobile keyboards
- [Input Method Editors](#)
- Predictive text and basic input aids
- Voice input where relevant
- AI-assisted text composition and correction tools

An analysis at this level would typically involve operating system input support, mobile platforms, IME availability, and observation of how users actually create text in real-world settings.

This is often the first point where users adapt themselves to the Internet rather than the Internet adapting to them.

User Interaction with Systems

Layer 4: Core Tools and Applications

For a language to be usable, it must function in the basic tools people rely on to exist online. This includes not just consuming content, but communicating, authenticating, and understanding system responses.

The question at this layer is not whether the experience is polished, but whether it is intelligible enough for users to act with confidence. Partial support here where some elements appear in-language, and others revert to a default, creates constant mental translation and quiet exclusion.

In scope of this layer:

- Web browsers including menus, dialogs, and settings
- Email clients including composition, reading, and system messages
- Messaging applications
- Authentication and account management flows
- System notifications and alerts

An analysis at this level would typically involve mainstream browsers, email clients, messaging platforms, and account systems, with attention to default language behavior and mixed-language flows.

This layer defines whether a language is merely present or actually usable.

Layer 5: Navigation and Interfaces

As users move through the Internet, they rely on interface elements to orient themselves and make decisions. Menus, settings, consent dialogs, error messages, and help text shape whether users understand what the system is asking of them and feel confident acting on it.

When interface language is inconsistent or missing, users guess. They click through without understanding. Over time, this erodes trust and agency. Inclusion at this layer is less about linguistic polish and more about reducing friction and fear.

In scope of this layer:

- Menus and navigation elements
- Settings and preferences
- Consent dialogs and notices
- Error messages and system feedback
- Help and support text

An analysis at this level would typically involve reviewing user interfaces across major platforms and services, especially points where users must make decisions or respond to system prompts.

Participation in the Broader Internet

Layer 6: Content Availability

Technical support without content leads to an empty experience. For a language to feel present, there must be something to engage with, such as information, education, culture, and community.

This does not require exhaustive coverage, but it does require sufficient breadth and relevance for users to feel that their language belongs online. People need to be able to publish information in their own language.

In scope of this layer:

- News and informational content
- Educational and reference material
- Cultural and community resources
- Public-interest information
- Locally relevant digital content

An analysis at this level would typically involve surveying major content platforms, content management systems (CMSs), public information sources, and community-driven spaces to understand what is actually available in-language.

Layer 7: Algorithmic Mediation and AI Systems

Even when content exists and is technically accessible, users increasingly encounter it through algorithmic systems. Search engines, recommendation systems, automated translation tools, generative AI assistants, and large language models now mediate how language is interpreted, summarized, surfaced, and understood.

This layer determines not only whether content can be found, but whether it is accurately represented, culturally understood, and usable. In many contexts, AI systems now replace direct navigation entirely, synthesizing answers rather than pointing to sources. Those AI systems are first trained on data that was found on the Internet.

When languages are underrepresented in training data, poorly modeled, or treated as low-confidence by AI systems, the result is subtle but powerful exclusion. Content may exist, but it is misinterpreted, deprioritized, summarized incorrectly, or ignored entirely. Over time, this makes a language feel digitally marginal, even if technical infrastructure is in place.

In scope of this layer:

- Search indexing and ranking
- Query understanding and intent matching
- Recommendation systems
- Machine translation quality
- Speech recognition and synthesis systems
- Large language model training data representation
- Generative AI summaries and assistant responses
- Automated content moderation models

An analysis at this level would typically involve:

- Testing in-language queries in search and AI assistants
- Evaluating translation and summarization fidelity
- Reviewing dataset representation in training corpora
- Assessing whether culturally specific knowledge is preserved or distorted
- Examining how AI confidence scoring affects ranking and surfacing

Failures at this layer signal that a language may be technically supported, but algorithmically diminished.

Layer 8: Transactions and Participation

Participation goes beyond reading and browsing. People need to be able to act by submitting forms, completing transactions, accessing services, and participating in civic and commercial life in their own language.

When languages disappear at moments of consequence, users are included only at the surface level. They are present, but not empowered.

In scope of this layer:

- Forms and submissions

- Payments and commerce
- Government and civic services
- Account changes and confirmations
- Meaningful online participation

An analysis at this level would typically involve service platforms, e-commerce systems, and public-sector interfaces where language directly affects the ability to act.

Layer 9: Trust, Safety, and Accountability

Finally, language must exist where risk, responsibility, and harm are involved. Warnings, security notices, moderation messages, and appeal processes determine whether users can protect themselves and understand their rights.

Absence at this layer does not just inconvenience users. It increases vulnerability.

In scope of this layer:

- Security and fraud warnings
- Scam and abuse notices
- Moderation and enforcement messages
- Appeals and dispute resolution
- Terms and policies that materially affect users

An analysis at this level would typically involve trust and safety systems, including automated and AI-assisted moderation workflows, and user protection mechanisms across platforms and services.

How This Draft Proposal Is Intended to Be Used

This draft proposal is not a checklist, ranking, or compliance tool. It is a diagnostic lens offered to the community.

By examining the Internet experience layer by layer, it becomes possible to see where a language drops out, where exclusion accumulates, and which gaps matter most for real users. Different languages will be strong in different layers and weak in others. The goal is not to score or compare languages, but to create a shared vocabulary for identifying structural barriers and prioritizing meaningful interventions.

We need to identify the gaps so that we as a global community can work to find the solutions so that ultimately every person can navigate the Internet in their own language.