

# *Halis Sefaradi*: critical authenticity and the voice of the first Ladino chatbot *Estreyika*

Alejandro Acero Ayuda (University of Oregon) & Carlos Yebra López (California State University, Fullerton)

aaceroay(at)uoregon.edu, cyebra Lopez(at)fullerton.edu

## Abstract

As the first Ladino-speaking chatbot intended to support language revitalization, *Estreyika* inhabits a complex ideological space where *authenticity* is constructed, contested, and operationalized through cultural, sociolinguistic, and political dynamics anchored in the lived realities of the last speakers of the endangered language Ladino: a Romance language of Ibero-Ottoman origin, preserved by the Sephardic diaspora for over 500 years, now considered severely endangered by UNESCO. In this short article, we argue that *Estreyika* functions as a site for negotiating performative ‘authenticity’, shaped by the learners’ positionalities—whether as heritage language learners, descendants of Ladino-speaking communities, or so-called ‘cultural outsiders’. Their linguistic performances become acts of alignment or resistance within a discursive field that privileges certain registers, geographies, or orthographies. By exploring *Estreyika* as a socio-political artifact, this paper advocates for an understanding of *authenticity* rooted in multiplicity, reflexivity, and critical positionality—a counter-narrative toward more inclusive LLMs critically attuned to AI-human partnerships.

## Keywords

AI-human interaction, LLM, authenticity, heritage speakers, endangered languages, Ladino

## 1 The endangerment of a living language

Ladino, also known as Judeo-Spanish, is a Romance language of Ibero-Ottoman origin (Papo 2020) historically spoken by the descendants of the Sephardic Jews expelled from the Iberian Peninsula in 1492. Today, the number of speakers has diminished and is mostly confined to older adults, intergenerational transmission is largely broken, and the language is classified by UNESCO (2003) as ‘severely endangered’.

The history of the Sephardic diaspora is marked by pivotal moments that in retrospect have been used to identify specific periods. Following his own framework to break down the Ashkenaz diaspora (1973), Max Weinreich subdivided the Sephardic diaspora in three broad stages (Sepharad 1-3), and scholars in the subject have echoed this subdivision (Hassán 1995; Hernández González 2001; Šmid 2002; Álvarez López 2017). Recently, Carlos Yebra López (2024) has extended this typology by incorporating a fourth critical point now that the global Ladino-speaking diaspora connects primarily online, resulting in a subdivision of three offline stages (Sepharad 1–3) plus one digital (Sepharad 4).



Acero Ayuda, Alejandro & Yebra López, Carlos. 2026.

*Halis Sefaradi*: first Ladino chatbot *Estreyika*

Special Issue: *The notion of authenticity*. Vol. 4 No.1

DOI: 10.62408/ai-ling.v4i1.54

*AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses*

ISSN: 2943-0070

CC-BY-NC-SA 4.0

Sepharad 1 refers to the Jewish presence in the Iberian Peninsula lasting from the 1st century CE until the Alhambra Decree of 1492 that marked their expulsion. During this period, proto-Ladino began to take shape in Romanized Iberia as Jewish immigrants interacted with local Romance varieties (Bunis 2019) already influenced by Iberian Arabic. By the thirteenth century, Ladino had also developed a parallel function as a written, word-for-word calque of Hebrew used for liturgical purposes (Séphiha 1977).

Sepharad 2 encompasses the period after the expulsion of the Sephardim in 1492 and their subsequent diaspora across the Mediterranean basin, lasting until the mid-20<sup>th</sup> century. In this phase, the 15th-century Iberian Romance varieties they brought to the diaspora, incorporated features from contact languages where they settled, mostly Turkish, Greek, South-Slavic, Italian or French, along with already present liturgical Hebrew (Minervini 2006). Over time, Ladino consolidated as a koine, a common language that developed organically due to close interaction among mutually intelligible Iberian Romance varieties spoken in the Mediterranean basin (Papo 2020; Yebra López 2021). During Sepharad 2, Ladino arises as a co-territorial diasporic vernacular with distinctive linguistic structures, a vibrant literary tradition, and tied to Sephardic cultural practices (Díaz-Mas 2006) with two principal variants: Haketia, western variety spoken in northern Morocco, and Judezmo, eastern variety prevalent across the Ottoman Empire.

Sepharad 3 designates the global diaspora of Sephardic communities beginning in the mid-twentieth century. This phase was propelled by historical ruptures, including the Holocaust (1933-1945), Turkey's wealth tax in the 1940s, the foundation of the State of Israel in 1948, and the Istanbul pogroms of 1955. At the turn of the 21<sup>st</sup> century, estimates suggested that there were fewer than 50,000 speakers worldwide (Ethnologue 2024) living in scattered communities across Turkey, Israel or the United States. The lack of spaces for vernacular use of Ladino, the decay of the intergenerational transmission, and the existing sociopolitical and ideological factors at play, forced many of these speakers to shift rapidly to the dominant languages where they primarily resided (Turkish, Hebrew, English or Spanish). In this stage, Ladino transitioned from a co-territorial vernacular rooted in distinct regions to a post-co-territorial diasporic vernacular in which geolectal distinctions were progressively disappearing, along with its Ladino-speaking community. In less than a century, Ladino shifted from being an "international trade language [...] in the Levant, to the colloquial language [...] for older people" (Harris 1994: 228).

## **2 Sepharad 4: Ladino Digital Renaissance**

The same COVID-19 pandemic that caused invaluable human losses and severe economic downturns also forced widespread societal changes. The seclusion imposed by the quarantine forced a shift in the modes of socialization, expediting a process of migration from the physical to the digital realm. For the Ladino-speaking community, this meant an opportunity to populate existing Web 1.0 and 2.0 platforms that were already invested in fostering worldwide community building (*Ladinokomunita*, 1999 – Rachel Amado Bortnick) or documenting the language and its speakers in the present (*Ladino21*, 2017 – Carlos Yebra López, Alejandro Acero Ayuda and Benni Aguado) but it also represented an unprecedented chance for creating new digital spaces where more innovative forms of intergenerational transmission can emerge, digital spaces where unrelated individuals of all generations are learning Ladino from and with each other, and co-constructing linguistic practices and ideologies together (e.g. *Los Ladinadores*, 2020 – Aldo Sevi; *Enkontros de Alhad*, 2020 – Liliana and Marcelo Benveniste, etc.).

The global diasporic community of Ladino speakers embraced with enthusiasm this virtual migration to a new Digital Home-Land (Held 2010; Yebra López 2021) where the identity of the Sephardi ‘imagined community’ (Anderson 1983) could be digitally articulated. The internet offered an online territory suited for the offline global diaspora to inhabit, recreating a virtual place for cohesion where Ladino was to be the lingua franca of the 21<sup>st</sup> century multilingual Sephardi. The Ladino speakers’ resourceful use of this technological innovation to imbue Ladino through a wide range of multimodal affordances on the digital world has led some Ladino activists to call this phenomenon the *digital renaissance* (Cruz Çilli 2021) of Ladino on the internet. The online articulation of the Sephardic diaspora in the 21st century is, according to Yebra López, what initiated Sepharad 4, and marked the reconfiguration of Ladino as a cyber-(post) vernacular (2024) language.

The online proliferation of manifold platforms and modes (WhatsApp and Facebook groups, Zoom gatherings, YouTube channels, email lists, Instagram, Snapchat or TikTok accounts, etc.) (see Yebra López 2024 for a detailed account) developed by the global, mainly digitally-native, community of Ladino speakers has revealed a new paradigm of sociolinguistic interest (Held 2010; Yebra López 2021; Bürki 2021) that is reshaping language and community in a dialogical fashion: the social construction of a digital Sephardi cultural identity (Lindholm 2013; Van Leeuwen 2001) that is redefining the expressive ‘authenticity’ of what counts as *halis* (pure, genuine, authentic) Ladino.

### 3 *Halis Ladino*: ‘Authenticity’ or the authentication of a language

On these digital platforms, heritage speakers (HS) of Ladino—those who acquired the language through home-based intergenerational transmission (Polinsky and Kagan 2007; Montrul 2016) coexist with second (L2) or most commonly multiple (L2+) languages learners, who encountered the language later in life through online community initiatives rather than within the family unit (King 2000). However, both groups now rely on these virtual spaces as their primary arenas for (post)vernacular expression and interaction. In this context, Ladino has become a contested digital space sustained partially by the nostalgia of an evoked ancestral code (Woodbury 2011) that is continuously emulated and prescribed, but also by the creativity of unique *linguaging* practices (García and Wei 2014) that generate new forms of authentication (Yebra López 2024).

These *linguaging* practices either adhere to the sociolinguistic boundaries and ideological expectations of Ladino as a static system or engage in the dynamic process of meaning making and knowledge shaping by deploying the speakers’ full linguistic repertoire without regard for purist constraints (Otheguy, García and Reid 2015). Regardless, their linguistic performances are not value-neutral; they are informed by the Ladino speakers’ positionalities—whether as heritage language learners, descendants of Ladino-speaking communities, or ‘cultural outsiders’ acquiring the language as part of their multilingual repertoire. Their performances become acts of alignment or resistance within a discursive field that privileges certain registers (formal, informal, liturgic, vernacular) (Séphiha 1977; Bunis 2021), geographies (Istanbul, Sarajevo, Salonica, Izmir, etc.) (Quintana 2007; Schwarzwald 2019), or orthographies (French-Ottoman Latin transcription, Rashi, Solitreo, etc.) (Bunis 2019; Yebra López 2024) over others.

In this context, the notion of *authenticity* cannot be disentangled from hegemonic assumptions about linguistic purity or correctness—both glottophagic processes (Calvet 1974), that is, sociopolitical dynamics within power-imbalanced linguistic environments that contribute to language shift (Fishman 1991) in an already endangered language. Instead of ‘authenticity’, Mary Bucholtz proposes the concept of *authentication* (2003) as an alternative term. ‘Authenticity’ is achieved through authenticating practices of those who use and evaluate language. Thus ‘authenticity’ is a socially negotiated process, rather than a fixed state (Bucholtz 2003). From a sociolinguistics perspective, as we have seen in Sepharad 2-3, it is problematic for Ladino prescriptivists to wield any variety as more “authentic” than others. Ladino is a pluricentric language with different geolects that developed in diaspora, around the various centers of economic, cultural, and political activity of the modern Sephardic diaspora such as Thessaloniki, Istanbul, Safed, Vienna, or Belgrade (Quintana 2010) and converged in the digital realm of Sepharad 4.

#### **4 *Estreyika*: AI-powered chatbot to revitalize Ladino**

The history of Sepharad 4 aligns with successive iterations of the Internet (Web 1.0, Web 2.0). The prospect of a more decentralized, intelligent, and autonomous Web 3.0 has already yielded technologies like blockchain, machine learning, and artificial intelligence. However, as previous iterations of the Web 1.0, 2.0 demonstrated, developments in the digital era tend to aggravate the digital divide between prestiged and minoritized languages (Kornai 2013). Trending digital tools like artificial intelligence (AI) build upon existing digital spaces and resources where prestiged languages are already fully established and with sufficient web-presence. Minoritized languages lack the material and human resources to power the creation and maintenance of such tools (Öktem, Zevallos, Moslem, Güneş et al. 2022). Limited textual resources, scarce state support, and historical patterns of societal and institutional marginalization hinder the participation of minoritized languages in the digital sphere (Nekoto, Marivate, Matsila et al. 2020). For that reason, endangered languages integration in these new Web 3.0 systems must be intentionally strengthened with adequate resources and policy support to back every community initiative by language activists, ideally by institutions (Tollefson and Pérez-Milans 2019). An example of a successful community initiative that benefited from institutional support with promising results was the project carried out by the *Col-lectivaT* and *Sephardic Center of Istanbul* that secured funding within the framework of the CCH-II (Ministry of Culture and Tourism of the Republic of Turkey with the financial support of the European Union.

Öktem, Zevallos, Moslem, Güneş et al. (2022) developed the first neural based system for Ladino capable of bidirectional machine translation and speech synthesis between Ladino - Spanish, Turkish - English, based on a baseline system of machine translation for Ladino. Drawing upon the potential of this advancement and relying on the last Natural Language Processing (NLP) and Machine Learning (ML) technology, Carlos Yebra López, Alejandro Acero Ayuda, and Alp Öktem created the first Ladino-speaking chatbot in history: *Estreyika* (2025).

#### **5 “It’s alive!”: from Chatbot to *Estreya Perez***

The idea for developing a Ladino-speaking chatbot originated with Carlos Yebra López, who envisioned the use of Web 3.0 technology to create an AI-powered digital agent capable of using natural language processing (NLP) to engage users in Ladino. At this incipient stage, Alejandro Acero Ayuda created the name, avatar, and narrative backstory that framed the chatbot’s persona, *Estreya Perez* (2025).

*Estreyika* was designed as a tool for language education and practice, equipped with the metalinguistic ability (Bialystok and Ryan 1985) to analyze the Ladino language and generate complex explanations of its linguistic phenomena, within the current limits of this promising field of study (Beguš, Dąbkowski and Rhodes 2025). At the same time, *Estreyika* was imagined as a narratively situated

character, embodied in a digital media capable of enacting a heightened ‘storyworld’ (Murray 1997) where multiple narratives are contained, and require the users’ active engagement to untap the innumerable possibilities of such immersive environment (Landow 2006). The backstory of Estreyika permeates through the storyworld that places the chatbot within a historically plausible context. Estreyika uses the most widespread, accepted variety of Ladino: the Istanbul gelect, written in the French-Ottoman Latin transcription according to the *Aki Yerushalayim* orthography, yet she understands/computes other Ladino gelects. The backstory of Estreyika is thus rooted in late eighteenth-century Istanbul, during a period of Sephardic cultural flourishing yet amid the broader transformation of the Ottoman Empire (Benbassa and Rodrigue 2000). The personal story embedded in Estreyika emerges in her replies to questions about daily life, dynamics of community, and conversations in Ladino. Interacting with her responses allows us to picture everyday life in a Jewish quarter of that era, where Ladino resonated in homes, streets, and markets, describing an environment where knowledge and learning is tied to the imagined social practices (Gee 2003) that Estreyika describes. Since its debut, Estreyika’s persona has been enthusiastically welcomed by the Ladino-speaking community.

The naming of the chatbot followed naturally from this vision of a chatbot companion to guide Ladino learners in their educational journey. Acero Ayuda reflected on the role of the chatbot not only as a pedagogical tool but also as a symbolic resource for the future of the language. The name, therefore, needed to evoke guidance, enlightenment, and orientation. Drawing inspiration from maritime navigation, where sailors relied on the stars to find their way home, he selected the name *Estreya* (Star) or the diminutive *Estreyika* (little star) to signify a light that shines, directs, and inspires future generations of Ladino learners committed to uphold the language’s revitalization.

After developing a detailed description of Estreyika’s backstory and personality, it was necessary to test whether the chatbot responded in ways consistent with her backstory, displaying coherence with the given identity and linguistic consistency as a key component to build user trust and engagement (Luger and Sellen 2016). To this end, Acero Ayuda designed a questionnaire organized around ten thematic areas: (1) identity, (2) synonyms and antonyms, (3) translation, (4) culturally significant topics, (5) proverbs, (6) grammar, (7) verb usage, (8) pronunciation, (9) historical knowledge and context, and (10) open-ended prompts that allowed Estreyika to engage in free conversation. Each theme was explored through multiple rounds of questioning, with increasing levels of detail requested in the responses. Following this iterative process, Acero Ayuda systematically recorded the strengths of Estreyika’s answers, identified errors requiring correction, and noted areas for improvement. This feedback was then provided to Alp Öktem who refined the system and enhanced Estreyika’s performance, resulting in the robust version available today.

The first prototype of *Estreyika* was tested in fall 2024 and launched in summer 2025 as an Alpha version, using Claude Sonnet 3.5 as the back-end model and Telegram as the front-end interface. Claude Sonnet belongs to the family of large language models (LLMs) developed by Anthropic, designed to generate human-like text through probability-based prediction across vast corpora of training data. While Claude Sonnet 3.5 allowed us to test *Estreyika*'s basic functionality, the Beta version—currently under development with Claude Sonnet 3.7-4.5—is designed to improve contextual sensitivity, reduce errors in low-resource linguistic environments, and expand multimodal affordances (development is scheduled to continue until September 2026). Telegram was chosen as the primary interface because of its global accessibility and user-friendly integration with Bot APIs (Application Programming Interface).

The challenges *Estreyika* faced were not merely technical but also structural. As Öktem, Zevallos, Moslem, Güneş et al. (2022) emphasized, marginalized languages such as Judeo-Spanish occupy an uneven position in the digitalized world. The lack of sufficient representation in digital corpora and institutional support required a creative, resource-conscious approach to train *Estreyika*. Instead of relying on massive corpora, the project drew upon curated sources of Ladino, including community-generated materials, archival texts, and prior digitalization efforts spearheaded by activists and scholars (e.g., Öktem's earlier work on building the first neural based system for Ladino capable of machine translation and speech synthesis).

*Estreyika* was further trained through iterative reinforcement provided by both heritage (HL) and multilingual (L2+) learners of Ladino, whose interactions helped refine *Estreyika*'s responses. This collaborative process highlighted both the opportunities and the limitations of working with endangered languages in the digital age: while human contributions can partially offset the lack of machine-readable data, they also foreground the extent to which “authentic” linguistic production is mediated by algorithmic constraints and ideological expectations within the speech community (Linnekin 1991).

The learning phase was finalized with the active participation of Ladino learners enrolled in California State University, Fullerton courses SPAN580T, SPAN468 (Fall 2024), SPAN468, and SPAN480T (Spring 2025) under the guidance of Carlos Yebra López. Their iterative interactions—ranging from testing conversational exchanges to evaluating lexical accuracy and stylistic register—significantly shaped *Estreyika*'s development. The chatbot's beta-phase refinement was supported by a Junior Research Grant, a Supercomputing Grant, and the research assistantship of Michelle Ramirez, providing critical resources for technical development and system optimization.

However, their voices were inevitably mediated both by the algorithmic structures of the large language model (LLM) and by the ideologically invested expectations of the remaining Ladino-speaking community. The Alpha phase revealed that *Estreyika* inhabits a complex ideological space where ‘authenticity’ is

constructed, contested, and operationalized through cultural, sociolinguistic, and political dynamics (Silverstein 1979; Irvine and Gal 2000; Blommaert 2005; Kroskrity 2010). Yet these processes remain anchored in the lived realities of the supposedly last speakers of Ladino, who carry the historical weight of the language's endangerment as well as the hopes of its revitalization. The participants' responses showed how Estreyika progressively evolved from a mere technical artifact, to become a self-referential discursive site in which both participants and Estreyika partook in the authentication process of Sephardic digital identity.

## **6 Student Perspectives on *Estreyika*: Legitimation through Authentication**

The students' interactions with Estreyika illuminated how learners themselves invoke notions of the “real”, the “authentic”, and the “true” to describe their experiences. Their testimonies bear witness of how ‘authenticity’ is taken at face value, embraced as an objective, self-evident property of speech or identity, rather than understood as a socially negotiated and contextually defined process of authentication (Creese, Blackledge and Takhi 2014). This dynamic is particularly relevant in the case of endangered languages such as Ladino, where legitimacy is often contested among the members of the diasporic community.

### **6.1 Feeling *Real* in Digital Encounters**

One of the most salient themes in the students' essays was the sense of realness that emerged from their exchanges with Estreyika. Several learners expressed surprise that the chatbot “felt like I was actually talking to a real person living in the region” (Student Essay). Crucially, students noted that Estreyika's delivery was often infused with warmth, nostalgia, and personality. Phrases such as “with a nostalgic smile” or “with bright eyes” accompanied their responses, lending the chatbot a deeply human quality. One student noted that interacting with Estreyika felt like “sitting across from a wise woman in the family,” a remark that blurred the line between bot and *nona* (grandmother), between mere transmission of information and intergenerational storytelling as a pedagogical strategy. These accounts suggest that ‘authenticity’, in this context, was not primarily linked to technical accuracy or linguistic purity, but to the affective impression of cultural resonance.

Such reactions underscore the relational dimension of ‘authenticity’: it emerges from interaction and recognition rather than from fixed linguistic criteria (Bucholtz 2003). Estreyika's evocations of Sephardic domestic life—markets, Turkish coffee, family gatherings—provided recognizable semiotic cues through which students authenticated her persona as “real.”

## 6.2 *Authentic* as Cultural Immersion

For many students, *authenticity* was framed in terms of immersion in cultural context. One learner noted: “Even if I could not verify every historical or culinary detail, the interaction felt like being immersed in an authentic cultural setting” (Student testimony). Another wrote that conversations about food were “a gateway into Ladino, making the interaction feel genuinely authentic”. *Estreyika* taught that to speak Ladino is not merely to master a vocabulary set, but to participate in a worldview: one where rituals are remembered, festivities are celebrated communally, and history is spoken out loud. Through everyday dialogue and richly detailed storytelling, *Estreyika* offered her users a rare experience: to learn a language not only about culture, but within it.

These responses reveal how ‘authenticity’ is produced through oral practice: talking about food, rituals, and cultural traditions served as semiotic resources that allowed learners to confer ‘authenticity’ on the chatbot. Moreover, this mode of authentication was not dependent on genealogical heritage but on discussed participation in symbolic practices that mediated belonging to the Sephardic ‘Digital Home-Land’ (Held 2010).

## 6.3 Negotiating *True Ladino*

Several students explicitly linked *authenticity* to language form, once again invoking terms such as “real” or “true.” One participant stated: “Having real conversations in a new language is, without doubt, the best way to learn” (Student testimony). Here, the notion of *real* was associated not with structural fidelity to a standardized Ladino but with the dialogic experience of conversational practice.

Other learners drew attention to specific linguistic features regarding orthography, such as the substitution of *k* for *c* or the alternation between *b* and *v* following the widely used, yet not exclusively, French-Ottoman orthography as prescribed by the magazine *Aki Yerushalayim*. This orthographic ascription was perceived as a marker of *linguistic authenticity*.

These discoveries reinforced their sense of engaging with the *true* form of Ladino, suggesting that for them, *authenticity* was less about fidelity to one historical register and more about the opportunity to inhabit Ladino as a living, learnable *lect*.

## 6.4 Authenticity, Diversity, and Community Anxiety

The testimonies also resonate with broader debates within the Ladino-speaking community. Brink-Danan (2011) has shown that ‘authenticity’ often functions as a gatekeeping device, distinguishing between *ladinoavlantes* and *no ladinoavlantes*, or disparaging ‘mixed’ varieties. Several students expressed moments of doubt or discomfort when *Estreyika*’s responses appeared inconsistent or too elaborate,

revealing the underlying anxiety tied to judgments of “authentic” use. Moreover, it is important to note that only a minority of participants in today’s digital Ladino fora consider Ladino their ‘mother tongue’ (Bunis 2020; Brink-Danan 2011), and that the community is increasingly heterogeneous (Santacruz 2019). Thus, privileging authenticity as defined by nativeness, genealogical descent, or unbroken intergenerational transmission risks excluding precisely those learners—heritage or L2—whose engagement is crucial for revitalization. The students’ emphasis on “authentic” or “true” experiences illustrates this tension: what they authenticate as real and meaningful may not align with community standards but nonetheless sustains their investment in Ladino.

### **6.5 Toward an Ethics of Authentication**

The students’ reflections ultimately demonstrate that ‘authenticity’ is less a property of linguistic forms than a dynamic sociolinguistic negotiation achieved through interaction, affect, and recognition (Avineri 2012). Their descriptions of Estreyika as “real,” “authentic,” and “true” confirm that ‘authenticity’ is always mediated by contextual practices. Thus, the question is not what is ‘authentic’, but rather what it means to be authentic here, under these norms, and “what are the authenticating practices by which it [authenticity] is conferred or denied?” (Creese, Blackledge and Takhi 2014: 939). Estreyika’s success, as reflected in student responses, lies in her capacity to enable authenticating practices—opportunities for learners to align with cultural traditions, rehearse identity, and imagine themselves as participants in the living Sephardic continuum of the 21<sup>st</sup> century. The *authenticity* they experienced resided less in inherited linguistic purity than in the affective and pedagogical outcomes of their engagement. However, does this pattern also hold for *ladinoavlantes* within the broader Sephardic ‘imagined community’?

## **7 Discussion: evaluations of Estreyika from *La Komunita***

To explore Ladino-speaking community perceptions of Estreya’s authenticity, we distributed a supplementary survey to a selected number of members of the Ladino online networks (N=10) with varying levels of self-reported proficiency, determined by a simplified LEAP-Questionnaire (Blumenfeld and Kaushanskaya 2020) included in the survey. Participants were prompted either to engage in unstructured interactions with Estreya or to rely on a bank of questions that mirrored the parameters and themes emerging from the students’ interactions: i) Feeling *Real* in Digital Encounters; ii) *Authentic* as Cultural Immersion; iii) Negotiating *True* Ladino; iv) Authenticity, Diversity, and Community Anxiety; and v) Toward an Ethics of Authentication. The survey questions elicited the same themes as those identified in the student interactions.

Of the 10 Ladino-speaking participants, 6 reported that Ladino was their L1, learned in the family since childhood; 2 reported learning Ladino as an L2 at a young age (under 18); and 2 reported learning Ladino as an L2 in adulthood (over 18). When inquired about their perceived ‘authenticity’ of *Estreya*, four out of ten respondents evaluated *Estreya*’s production as generally or “fully authentic”, two respondents argued that the chatbot was “both authentic and inauthentic” concerning different aspects, and one participant incurred in item nonresponse (missing this specific question). Three respondents expressed reservations, citing issues such as Castilianisms, lexical and orthographic inconsistencies, a limited stylistic range, and the use of overly “elitist” vocabulary. Importantly, even those who did not fully accept *Estreya* as “authentic” often described moments in which the chatbot *felt* real, whether through emotional responsiveness, conversational alignment, or an impression of interpersonal attentiveness (the ability to *sense* and respond to the user’s input).

Moreover, 3 of the 4 respondents who evaluated *Estreya* as “authentic” self-reported in the background data section that they were L1 speakers of Ladino, “learned in the family since childhood”. The fourth respondent reported being an L2 speaker of Ladino that “learned the language at a young age (under 18)”. This distribution reinforces the argument that ‘authenticity’ is not an inherent property of linguistic forms, (as these 4 respondents also included mention of linguistic features in their critiques) but a dynamic sociolinguistic negotiation achieved through interaction, affect, and recognition (Bucholtz 2003). The evaluations ranging from detailed linguistic critique yet emphasizing the affective identification, demonstrate that perceptions of ‘authenticity’ depend on the relational experience and not solely on strict adherence to linguistic norms or forms, echoing the positive tendencies observed among the CSU Fullerton learners in the initial testing phase.

We acknowledged the limitations of our student sample, including existing power dynamics between instructor and students, and participants’ limited prior knowledge of Sephardic culture. While triangulation with L1, L2, L2+ Ladino speakers from the broader digital community strengthens our findings, limitations remain, including the small size of the Ladino-speaking community sample (N=10), potential self-selection bias, and the exclusively digital nature of the exchange, which may have affected representation of less digitally engaged community members. Future research could examine a larger pool of participants representing greater diversity in demographics, Ladino proficiency levels, and degrees of engagement with the online Sephardic diaspora from ‘Sepharad 4’ (Yebra López 2024).

## 8 Conclusions

The *Estreyika* model, while pioneering in its innovative approach within the Ladino-speaking world, remains bounded by several unresolved questions and

methodological caveats. A central issue is the uncertainty surrounding whether large language models genuinely internalize human-like, offline cognitive patterns and biases, thus reproducing ‘algorithms of oppression’ (Noble 2018) or merely reflect statistical regularities of language use. This ambiguity complicates efforts to assess the model’s reliability and interpretive depth when applied to minoritized languages such as Ladino. Furthermore, the pragmatic decision to privilege the Istanbul geolect in French-Ottoman Latin orthography—though it maximizes accessibility and standardization—comes at the cost of constraining the broader landscape of Ladino’s diverse geolects, sociolects, and scripts.

Despite the current Alpha version accepting and computing lexical variety, it is crucial that we also maintain educational engagement with Rashi and Solitreo scripts, not only to preserve historical knowledge but also to sustain the cultural richness embedded in traditional writing systems.

The Estreyika model has opened the debate for the Ladino-speaking community, highlighting the dual task of language revitalization in the digital era: ensuring communicative viability while simultaneously reshaping ideological valuations of minoritized languages. Estreyika has already done more than simply provide access to pedagogical and linguistic resources; it has situated Ladino as a contemporary, future-oriented, revalued language (Eisenlohr 2004) that is integral to modern collective Sephardi identity.

The challenge moving forward will be to negotiate a balance between communicative efficiency and linguistic diversity in order to secure a revitalization of Ladino that is both sustainable and representative of the Sephardic idiosyncrasy. This process must also empower Ladino-speaking communities as active agents in shaping revitalization strategies, aligning with participatory models of language ownership that affirm local self-determination and ensure long-term sustainability (Kono 2019).

## Acknowledgments

The authors would like to *rengrasyar de las onduras de nuestro korason* to the Ladino-speaking community members for their enthusiasm, generous contributions and unwavering support throughout the development, testing and review of *Estreyika*. Special appreciation is extended to Alp Öktem, whose technical expertise and collaboration were instrumental in bringing the chatbot to life, and the Spanish students at California State University, Fullerton, whose active engagement provided invaluable feedback during the chatbot’s design and refinement. We also acknowledge the support of the CSUF Junior Research Grant, the Supercomputing Grant, and the research assistantship of Michelle Ramirez, without which this chatbot and article would not have been possible. Finally, we sincerely appreciate the reviewers’ valuable comments and suggestions, which helped us to improve the quality of the manuscript.

## Conflicts of interest

The authors declare no conflicts of interest regarding the publication of this contribution.

## References

- Álvarez López, Cristóbal. 2017. *Estudio lingüístico del judeoespañol en la revista 'Aki Yerushalayim'*. Ph.D. Dissertation. University of Seville. <http://hdl.handle.net/11441/68487> (last accessed on 16/02/2026).
- Anderson, Benedict. 1983. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.
- Avineri, Netta Rose. 2012. *Heritage Language Socialization Practices in Secular Yiddish Educational Contexts: The Creation of a Metalinguistic Community*. Ph.D. Dissertation. University of California. <https://escholarship.org/uc/item/9f50n171> (last accessed on 16/02/2026).
- Benbassa, Esther & Rodrigue, Aron. 2000. *Sephardi Jewry: A history of the Judeo-Spanish community, 14th–20th centuries*. Berkeley: University of California Press.
- Beguš, Gašper & Dąbkowski, Maksymilian & Rhodes, Ryan. 2025. Large Linguistic Models: investigating LLMs' metalinguistic abilities. *IEEE Transactions on Artificial Intelligence* 6(12). 3453–3467. <https://doi.org/10.48550/arXiv.2305.00948>
- Bialystok, Ellen & Ryan, Ellen B. 1985. Toward a definition of metalinguistic skill. *Merrill-Palmer Quarterly* 31(3). 229–251.
- Blommaert, Jan. 2005. *Discourse: A Critical Introduction*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511610295>
- Brink-Danan, Marcy. 2011. The meaning of Ladino: the semiotics of an online speech community. *Language & Communication* 31(2). 107–118. <https://doi.org/10.1016/j.langcom.2010.08.003>
- Bucholtz, Mary. 2003. Sociolinguistic nostalgia and the authentication of identity. *Journal of Sociolinguistics* 7(3). 398–416. <https://doi.org/10.1111/1467-9481.00232>
- Bunis, David. 2019. La Ortografía de Aki Yerushalayim: un pinakolo en la estoria de la romanización del djudezmo (djudeo-espanyol). *Aki Yerushalayim* 101. 8–24.
- Bunis, David. 2020. Sephardic customs as a discourse topic in the *Ladinokomunita* internet correspondence circle. In Lifshitz, Joseph Isaac & Feuchtwanger-Sarig, Naomi & S. Baumgarten, Goldin, & Diner, Hasia (eds), *Minhagim: Custom and practice in Jewish life*, 161–197. Berlin: De Gruyter.

- Bunis, David. 2021. La lingua de los sefaradim en romanizacion. In Schwarzwald, Ora (ed.), *Las Ortografías del Ladino*, 21–37. Jerusalem: La Akademia Nasionala del Ladino en Israel.
- Bürki, Yvette. 2021. El activismo lingüístico y la lengua sefardí en el siglo XXI. In Jovanovic, Zeljko & Sánchez Pérez, María (eds), *Ovras son onores: estudios sefardies en homenaje a Paloma Díaz-Mas*, 211–32. Bilbao: University of the Basque Country.
- Calvet, Louis-Jean. 1974. *Linguistique et colonialisme : petit traité de glottophagie*. Paris: Payot.
- Creese, Angela & Blackledge, Adrian & Takhi, Jaspreet Kaur. 2014. The ideal “native speaker” teacher: negotiating authenticity in a multilingual classroom. *The Modern Language Journal* 98(4). 937–951. <https://doi.org/10.1111/j.1540-4781.2014.12126.x>
- Cruz Çilli, Kenan. 2021. Ladino’s Renaissance: for this dying Jewish language, COVID has been an unlikely lifesaver. *Haaretz*, January 05, 2021. <https://www.haaretz.com/middle-east-news/2021-01-05/ty-article-opinion/.premium/ladinos-renaissance-covid-the-unlikely-lifesaver-for-a-dying-language/0000017f-f461-d47e-a37f-fd7df1120000> (last accessed on 16/02/2026).
- Díaz-Mas, Paloma. 2006. *Los sefardíes: Historia, lengua y cultura*. Riopiedras.
- Eisenlohr, Patrick. 2004. Language revitalization and new technologies: cultures of electronic mediation and the refiguring of communities. *Annual Review of Anthropology* 33. 21–45. <https://doi.org/10.1146/annurev.anthro.33.070203.143900>
- Enkontros de Alhad. 2020. Online weekly program in Ladino created by Liliana and Marcelo Benveniste. <https://esefarad.com/tag/enkontros-de-alhad/> (last accessed on 15/03/2026).
- Estreya Perez. 2025. Ladino language chatbot created by Yebra López, Carlos & Acero Ayuda, Alejandro & Öktem, Alp. <https://t.me/estreyaladinobot> (last accessed on 16/02/2026).
- Ethnologue. 2024. *Judeo-Spanish*. SIL International. Retrieved from <https://www.ethnologue.com/language/lad> (last accessed on 16/02/2026).
- Fishman, Joshua. A. 1991. *Reversing Language shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages*. Clevedon: Multilingual Matters 76.
- García, Ofelia & Wei, Li. 2014. *Translanguaging: Language, Bilingualism and Education*. New York: Palgrave Macmillan. <https://doi.org/10.1057/9781137385765>

- Gee, James Paul. 2003. What video games have to teach us about learning and literacy. *Computers in Entertainment* (1)1. 1–4. New York: Palgrave Macmillan.  
<https://doi.org/10.1145/950566.95059>
- Harris, Tracy. 1994. *Death of a Language: The History of Judeo-Spanish*. Newark, NJ: University of Delaware Press.
- Hassán, Iacob. 1995. El español sefardí (judeo-español, ladino. In Seco, Manuel & Salvador, Gregorio (eds), *La lengua española hoy*, 117–40. Madrid: Fundación Juan March.
- Held, Michal. 2010. The people who almost forgot: Judeo-Spanish web-based interactions as a digital home-land. *El Presente: Studies in Sephardic Culture* 4. 83–101.
- Hernández González, Carmen. 2001. Un viaje por Sefarad: la fortuna del judeoespañol. *El español en el mundo*. Instituto Cervantes/Plaza & Janés/Círculo de Lectores.
- Irvine, Judith T. & Gal, Susan. 2000. Language ideology and linguistic differentiation. In Kroskrity, Paul V. (ed.), *Regimes of Language: Ideologies, Politics, and Identities*, 35–84. Santa Fe: School of American Research Press.
- Kaushanskaya, Margarita & Blumenfeld, Henrike & Marian, Viorica. 2020. The language experience and proficiency questionnaire (LEAP-Q): ten years later. *Bilingualism: Language and Cognition* 22(2). 1–6.  
<https://doi.org/10.1017/S1366728919000038>
- King, Kendall. A. 2000. Language Ideologies and Heritage Language Education. *International Journal of Bilingual Education and Bilingualism* 3(3). 167–184.  
<https://doi.org/10.1080/13670050008667705>
- Kono, Nariyo. 2019. Linguistics: community-based language revitalization. In Iphofen, R. (ed.), *Handbook of Research Ethics and Scientific Integrity*, 1–13. Cham: Springer. [https://doi.org/10.1007/978-3-319-76040-7\\_45-1](https://doi.org/10.1007/978-3-319-76040-7_45-1)
- Kornai, András. 2013. Digital language death. *PLoS ONE* 8(10). 1–11.  
<https://doi.org/10.1371/journal.pone.0077056>
- Kroskrity, Paul. V. 2010. Language ideologies—evolving perspectives. In Jaspers, Jürgen & Östman, Jan-Ola & Verschueren, Jef (eds), *Society and language use*, 192–211. Amsterdam: John Benjamins. <https://doi.org/10.1075/hoph.7.13kro>
- Ladino 21. 2017. Community Interest Company. n.d. Ladino 21. [www.ladino21.org](http://www.ladino21.org) (last accessed on 16/02/2026).
- Ladinokomunita. 1999. Ladino language listserv created by Rachel Amado Bortnick. <https://ladinokomunita.groups.io/> (last accessed on 16/02/2026).
- Landow, George P. 2006. *Hypertext 3.0. Critical Theory and New Media in an Era of Globalization*. Baltimore: Johns Hopkins University Press.

- Lindholm, Charles. 2013. *Culture and Authenticity*. Malden, MA: Wiley-Blackwell.
- Linnekin, Jocelyn. 1991. Cultural invention and the dilemma of authenticity. *American Anthropologist* 93(2). 446–449.  
<https://doi.org/10.1525/aa.1991.93.2.02a00120>
- Los Ladinadores. 2020. Facebook group in Ladino created by Aldo Sevi. <https://www.facebook.com/groups/ladinadores/> (last accessed on 16/02/2026)
- Luger, Eva & Sellen, Abigail. 2016. “Like having a really bad PA”: The gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, May 7-12, 2016), 5286–5297. New York: Association for Computing Machinery. <https://doi.org/10.1145/2858036.2858288>
- Minervini, Laura. 2006. El desarrollo histórico del judeoespañol. *Revista Internacional de Lingüística Iberoamericana* 8. 13–34.
- Montrul, Silvina. 2015. *The Acquisition of Heritage Languages*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139030502>
- Murray, Janet H. 1997. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. New York: Free Press.
- Nekoto, Wilhelmina & Vukosi, Marivate, & Tshinondiwa, Matsila & Fasubaa, Timi & Fagbohunge, Taiwo & Akinola, Solomon Oluwole & Muhammad, Shamsuddeen & Kabenamualu, Salomon Kabongo & Osei, Salomey & Sackey, Freshia et al. 2020. Participatory research for low-resourced machine translation: a case study in African languages. In Cohn, Trevor & He, Yulan & Liu, Yang (eds), *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2144–2160. Association for Computational Linguistics.  
<https://doi.org/10.18653/v1/2020.findings-emnlp.195>
- Noble, Safiya. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.  
<https://doi.org/10.2307/j.ctt1pwt9w5>
- Otheguy, Ricardo & García, Ofelia & Reid, Wallis. 2015. Clarifying translanguaging and deconstructing named languages: a perspective from Linguistics. *Applied Linguistics Review* 6(3). 281–307.  
<https://doi.org/10.1515/applirev-2015-0014>
- Öktem, Alp & Zevallos, Rodolfo & Moslem, Yasmin & Güneş Öztürk, Özgür & Gerson Şarhon, Karen. 2022. Preparing an endangered language for the digital age: The Case of Judeo-Spanish. In Ojha, Atul Kr. & Liu, Chao-Hong & McCrae, John P. (eds), *Proceedings of the Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-resourced Languages in Eurasia within the 13th Language Resources and Evaluation Conference* (Marseille, June 20-25, 2022), 105–110. European Language Resources

- Association. <https://aclanthology.org/2022.eurali-1.18/> (last accessed on 16/02/2026).
- Papo, Eliezer. 2020. Kualo es el Ladino? Eliezer Papo i Alice Corr para Ladino 21 // Londra, Inglaterra. in Ladino 21. Published on February 7, 2020. YouTube video, 8:40. <https://www.youtube.com/watch?v=Fsb7xqSXGrU> (last accessed on 16/02/2026).
- Polinsky, Maria & Kagan, Olga. 2007. Heritage languages: in the ‘wild’ and in the classroom. *Linguistics and Language Compass* 1(5). 368–395. <https://doi.org/10.1111/j.1749-818x.2007.00022.x>
- Quintana Rodríguez, Aldina. 2007. *Geografía lingüística del Judeoespañol: estudio sincrónico y diacrónico*. Lausanne, Switzerland: Peter Lang. <https://doi.org/10.3726/978-3-0351-0275-8>
- Quintana Rodríguez, Aldina. 2010. El judeoespañol, una lengua pluricéntrica al margen del español. In Díaz-Mas, Paloma & Sánchez Pérez, María (eds), *Los sefardíes ante los retos del mundo contemporáneo: identidad y mentalidades*, 31–52. Madrid: Consejo Superior de Investigaciones Científicas.
- Santacruz, Daniel. 2019. Ladinokomunita: 20 Years Old and Still Going Strong. *KolSefardim: News and Features about the Sephardic World*, December 8, 2019. <https://www.kolsekardim.net/post/ladinokomunita-20-years-old-and-still-going-strong> (last accessed on 16/02/2026).
- Schwarzwald, Ora. 2019. Judeo-Spanish throughout the Sephardic Diaspora. In Benor, Sarah (ed.), *Languages in Jewish Communities, Past and Present*, 145–84. Berlin, Boston: De Gruyter. <https://doi.org/10.1515/9781501504631-007>
- Séphiha, Haïm Vidal. 1977. Archaïsmes lexicaux du ladino (judéoespagnol Calque). *Cehm Cahiers De Linguistique Hispanique Médiévale* 2(1). 253–61.
- Silverstein, Michael. 1979. Language structure and linguistic ideology. In Clyne, Paul R. & Hanks, William F. & Hofbauer, Carol L. (eds), *The Elements: A Parasession on Linguistic Units and Levels*, 193–247. Chicago: Chicago Linguistic Society.
- Šmid, Katja. 2002. Los problemas del estudio de la lengua sefardí. *Verba Hispanica* 10(1). 113–24. <https://doi.org/10.4312/vh.10.1.113-124>
- Tollefson, James W. & Pérez-Milans, Miguel. 2019. (eds), *The Oxford Handbook of Language Policy and Planning*. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190458898.001.0001>
- UNESCO Ad Hoc Expert Group on Endangered Languages. 2003. Language Vitality and Endangerment. UNESCO. <https://ich.unesco.org/doc/src/00120-EN.pdf> (last accessed on 16/02/2026).
- Van Leeuwen, Theo. 2001. *Discourse and practice: New tools for critical discourse analysis*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195323306.001.0001>

- Weinreich, Max. 1973 [2008]. Yiddish in the Framework of Other Jewish Languages. In Glasser, Paul (ed.), *History of the Yiddish Language*, 124–74. New Haven/London: Yale University Press.  
<https://doi.org/10.1017/S0037677900015242>
- Woodbury, Anthony C. 2011. Language documentation. In Austin, Peter K. & Sallabank, Julia (eds), *The Cambridge Handbook of Endangered Languages*, 159–186. Cambridge: Cambridge University Press.  
<https://doi.org/10.1017/CBO9780511975981.009>
- Yebra López, Carlos. 2021. The digital (de)territorialization of Ladino in the twenty-first century. *WORD* 67(1). 94–116.  
<http://doi.org/10.1080/00437956.2021.1880065>
- Yebra López, Carlos. 2024. *Ladino on the Internet: Sepharad 4*. London: Routledge. <https://doi.org/10.4324/9781003473664>