Journal of English Language Teaching and Applied Linguistics
ISSN: 2707-756X
DOI: 10.32996/jeltal
Journal Homepage: www.al-kindipublisher.com/index.php/jeltal

JELTAL

AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

From Analysis to Generation: Semiotic Prompt Engineering for AI-Generated Educational Visuals

Kettani Moulay Tahar and Ouahidi Lalla Meriem
English Department, Sultan Moulay Sliman University, Beni Mellal, Morocco
English Department, Sultan Moulay Sliman University, Beni Mellal, Morocco
Corresponding Author: Kettani Moulay Tahar; **E-mail**: Tahar.flsh.phd@usms.ac.ma

| ABSTRACT

Engaging students and developing intercultural literacy are two key benefits of using Educational Visuals (EVs); however, the selection of an EV is often driven by convenience rather than sound methodological processes. This research moves beyond the analysis of existing representations of Moroccan culture through the creation of a new Semiotic Prompt Engineering (SPE) framework. The current study uses Gunther Kress and Theo van Leeuwen's multimodal discourse analysis framework as a lens within the generative reality of AI models to examine the systematic curation of EVs representing educational effectiveness and cultural relevance. The proposed methodology consists of feeding an AI model a particular taxonomy and parameters that correspond to three metafunctions relative to representation, interaction, and composition, and shifting from static and independent representations to dynamic and narratively rich ones. The findings of this conceptual proof-of-concept with illustrative comparative renderings suggest that visuals generated through semiotically informed prompting techniques may portray greater depth of narrative and offer a more culturally grounded representation than visuals produced through conventional prompting techniques. In addition, the research invites a re-conceptualization of the roles of educators and curriculum designers from active collectors of EVs to knowledgeable and creative EV content creators.

| KEYWORDS

Educational visuals, multimodal discourse analysis, prompt engineering, semiotic prompt engineering, visual semiotics

| ARTICLE INFORMATION

ACCEPTED: 19 September 2026 **PUBLISHED:** 15 November 2026 **DOI:** 10.32996/jeltal.2026.8.11.5

1. Introduction

Visuals play a vital role in enhancing comprehension, stimulating curiosity, and encouraging intercultural understanding among both teachers and students. Therefore, in multicultural settings like Morocco's EFL classroom, it is important to curate visuals that depict cultural representations of all students and provide meaningful learning experiences that foster identity construction processes. Currently, there is a scarcity of inclusive and quality visual materials, as their selection process is often based only on what's readily available rather than what has educational potential or culturally complex narratives. This may produce visuals devoid of diverse and layered narratives and symbolism that give students an opportunity for in-depth intercultural engagement.

This research actively generates pedagogically informed visual materials with the use of generative AI rather than analyzing existing materials. Drawing on Gunther Kress and Theo van Leeuwen's (2006) social semiotic theory, this research explores how visual representations, interactive processes, and compositional organization can be operationalized through "Semiotic Prompt Engineering" (SPE), a methodology that operationalizes established semiotic principles as a structured grammar for directing AI image generation. Using the Nano Banana image generator in the Gemini application, this study explores how quality generative

Copyright: © 2026 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (https://creativecommons.org/licenses/by/4.0/). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

AI-generated visuals, grounded in purposeful semiotic principles and with appropriate methodologies, can ultimately serve as high-quality instruments for both contextualized and meaningful pedagogical practices.

This investigation seeks to find out whether Kress and van Leeuwen's (2006) three metafunctions of visual social semiotics, once applied systematically as a prompting strategy, are able to produce AI-generated educational visuals that show greater narrative depth, cultural specificity, and semiotic organization than those produced by conventional natural language prompting. To explore this question, the Moroccan "Tbourida", also known as the gunpowder game, was chosen as a case study. The Tbourida was inscribed on UNESCO's representative list of the intangible cultural heritage of humanity in 2021 (UNESCO, 2021). It should be stressed that the study is positioned as a proof of concept and not as a controlled empirical experiment.

2. Literature Review

The following review situates the SPE framework within three intersecting bodies of scholarship: visual social semiotics and multimodal discourse analysis (MDA), AI-generated imagery in education, and prompt engineering as pedagogical practice.

2.1 Visual Social Semiotics and Multimodal Discourse Analysis

Kress and van Leeuwen's (2006) theoretical framework for social semiotics as outlined in their book Reading Images: The Grammar of Visual Design was adopted for this study. Their framework adapts Halliday's (1978) theory of systemic functional linguistics and proposes that visual images can be understood on three levels: the representational, the interactive, and the compositional. According to Kress and van Leeuwen, people will make meaning from an image based on how that image represents reality (the representational function), how it interacts with the viewer (the interactive function), and how all elements within the image relate to each other (the compositional function). This tripartite framework is now widely used in MDA to analyze how visual images represent instruction in textbooks, digital technologies, and classroom materials (Abbas et al., 2026; Liu et al., 2024). Studies based on Kress and van Leeuwen's tripartite framework indicate that visual forms of representation available for use by teachers are never ideologically neutral; rather, they encode cultural values, represent social relationships, and influence learners' understandings of the validity of differing social experiences represented in the curriculum (Jewitt, 2009; Machin, 2007). In the case of EFL textbooks, Weninger and Kiss (2013) have argued that images and texts produce cultural meanings through a process of semiosis rather than containing culture as a fixed object, and that even images presented as cultural are often used mainly to build linguistic competence.

What the field has not yet developed, however, is a methodology for moving from analysis to generation, using semiotic principles not merely to critique existing visuals, but to actively produce new ones. This paper addresses that gap directly.

2.2 AI-Generated Imagery in Educational Contexts

Globally, the growing trend in the development of generative AI as well as Assistive Technology for Education has been focused on the automation of instructional material generation (Qian, 2025). While advances in AI offer personalized learning experiences, there appears to be a significant gap in methodological processes when developing culturally responsive personalized learning. This gap has become an area of concern for many authors, and recent research has shown that generative AI tends to reproduce dominant norms at the expense of linguistic and cultural diversity (Albeihi & Rice, 2025), which suggests that culturally responsive personalized learning requires a clearer framework to ensure that AI-generated imagery does not depict narratives lacking in depth and cultural context, thus allowing for the development of engaged learning opportunities. As such, the authors have argued that if AI is not supported by scientific and methodological rigor, the imagery generated may replicate the same illustrative convenience found in traditional textbooks, rather than produce positive pedagogical outcomes from the use of the AI-generated imagery (Bozkurt et al., 2024). This concern is supported by research on text-to-image models, which has found that such models tend to amplify demographic stereotypes on a large scale (Bianchi et al., 2023) and may reproduce an outsider's gaze toward non-Western cultures (Qadri et al., 2023).

2.3 AI and Representation in the Moroccan Educational Context

From a national perspective, generative AI is beginning to transform learning and teaching in the Moroccan educational system. Within this transformed education system, there is a growing need to implement adaptive learning tools to develop localized content for the diversity of Moroccan multicultural classrooms. Furthermore, scholars conducting research on Moroccan Amazigh documentary film have articulated that representational sovereignty must be maintained to ensure that visual

representations accurately represent the specific and unique cultural characteristics of a group (Petty & Benbouazza, 2020). This principle applies equally to visuals used as instructional materials if they are to be truly inclusive. Similarly, Ofosu-Asare (2025) has argued that for artificial intelligence to avoid reproducing cultural bias, it must be developed using a framework that acknowledges the particular epistemological and socio-cultural origins of a particular culture.

2.4 Prompt Engineering as Pedagogical Practice

Despite the aforementioned literature, few researchers have developed definitions and descriptions outlining how to operationalize Kress and van Leeuwen's (2006) three metafunctions as a direct prompting grammar for AI image generation models. The practice of prompt engineering, or crafting inputs to AI systems in ways that systematically shape their outputs, has emerged rapidly as a significant area of practical and theoretical interest, with most existing work focused on text-based language models (Brown et al., 2020; Wei et al., 2022). Research on text-to-image prompting has started to address this imbalance, with design guidelines identifying the prompt elements that most influence generated images (Liu & Chilton, 2022), taxonomies of the prompt modifiers used by practitioners (Oppenlaender, 2024), and structured prompting frameworks developed for Google's Gemini image models (Cazzaniga, 2026). However, these works approach prompting in terms of its practical effects rather than its meaning-making potential. Work that grounds visual prompting in an explicit theoretical framework for meaning-making is, to the best of the authors' knowledge, scarce and limited. As discussed in Section 2.1, MDA has been widely applied to existing instructional materials such as textbooks, and in Morocco, researchers have more recently used semiotic analysis to evaluate how AI interprets Moroccan cultural artifacts (El Farouki & Bouziane, 2025). However, to the best of the authors' knowledge, no published research has demonstrated how generative "Semiotic Prompt Engineering" can be used to create AI imagery; therefore, this points to a gap in the research to support the need for developing narrative-rich and layered AI-generated imagery. By using the three metafunctions to support the research question, this current research proposes a practical checklist (Appendix A) and a set of criteria for the development of educationally focused generative AI.

3. Research Methodology

The methodological approach used in this research to explore the intersection of visual social semiotics (VSS) with AI-powered image generation is outlined in this section. This research explores whether the structured semiotic framework based on the framework of Kress and van Leeuwen (2006) can be operationalized as a prompt engineering strategy to develop culturally and hierarchically nuanced renderings of the Moroccan Tbourida, known in English as the gunpowder game.

3.1 Research Instrument: Gemini and Nano Banana

The images were generated through the Images feature of the Google Gemini application (Google, 2026), which offers Google's Nano Banana image-generation capability (Google DeepMind, 2026). Gemini 4 Argon was selected in the application, with the thinking level set to High; however, the exact Nano Banana variant responsible for rendering the images was not recorded or independently verified. The generative system is therefore described as Nano Banana within Gemini rather than attributing image generation directly to Argon. Further research is needed to determine whether these results generalize to other image generators.

3.2 Research Design: Semiotic Prompt Engineering (SPE)

The research used a qualitative comparative approach to look at the differences between a natural language prompt and a longer SPE prompt for the same task, with three renderings per prompt condition. This examination is not a controlled experiment. Instead, it provides a proof of concept that illustrates how semiotic scaffolding may impact AI-generated visual output. Thus, the study aims at developing a process that can be repeated in future empirical research.

In this sense, the Tbourida case study is to be considered a proof of concept, which shows that SPE, as opposed to unstructured prompts, may produce visually rich narratives with greater cultural specificity. Ultimately, the study seeks to encourage further empirical research across different AI models and in a wider range of cultural, visual, and semiotic settings.

As for the baseline prompt, it was designed to resemble the informed, yet unsystematic, description that an educator would write on his or her own. In plain language, it asks for a realistic 4:3 image of the Tbourida that could be used with students in an English-language textbook, and it describes what the picture should contain: a Sorba of an odd number of riders, the Mqedem riding in the middle of the line rather than far out in front, the horses, the spectators standing at a safe distance, the tents, and a

bright, clear day; however, it makes no reference to semiotic theory, and it leaves the composition of the image (angle, distance, placement, salience) to the model. It opens as follows: "Create a horizontal 4:3 image of Tbourida that I can use as a teaching aid for my students in English classes." The full prompt (479 words, counted by whitespace-separated tokens) is reproduced in Appendix B.

The semiotically engineered prompt was constructed based upon the meta-functional mappings proposed by Kress and van Leeuwen (2006), using the tiered framework presented in Section 3.3. It asks for the same event, the same setting, and the same participants as the baseline prompt, including the odd-numbered Sorba and the position of the Mqedem inside the line; however, its instructions are organized in three sections, one per metafunction, and worded in the terms of the framework (Actors and vectors, demand and offer, center and margin, Ideal and Real, framing). The baseline prompt contains 479 words, while the SPE prompt contains 608 words (whitespace-separated count). Therefore, the two prompts differ in length as well as in the organization and specificity of their instructions; this comparison cannot isolate the effects of semiotic structure from those of prompt length or content. The SPE prompt is reproduced in full in Appendix B.

The images were generated through the Gemini application (Google, 2026), each in a fresh chat session, with each prompt submitted in full as a single prompt: the baseline prompt was run three times on 18 October 2026, and the SPE prompt three times on the same day. Both prompts ask for a 4:3 frame, and all the renderings came out at 2400 × 1792 pixels. The first rendering of each condition is analyzed in detail in Section 4 (Figure 1); the other renderings are shown in Figure 2 and are referred to whenever they confirm or qualify what is observed in Figure 1.

3.3 The Operational Framework

The methodology of this research focused mainly on building a tiered prompting framework grounded in Kress and van Leeuwen's (2006) visual social semiotics. This framework was used by the researcher to synthesize the experimental prompt, in a way that ensured that each visual aspect served a specific representational, interactive, or compositional objective. The output of each tier became one section of the SPE prompt reproduced in Appendix B, under the heading of its metafunction. The instructions derived from each metafunction are summarized in Table 1.

Since the study is a proof of concept, the framework should not be taken as a finalized or validated instrument; it is rather offered as a model that is open to critique, refinement, and replication in other cultural contexts.

Table 1

SPE Instructions Used in the Study, by Metafunction

Metafunction	Parameter	Instruction in the SPE prompt
Representational	Action	Riders and horses as Actors; convergent forward vectors; one instant just after the collective shot, barrels skyward, smoke emerging simultaneously; coordination, not isolated shooting
	Reaction	Spectators' gestures and gazes directed toward the Sorba
	Speech	Mqedem's commanding gesture and recently shouted signal
	Circumstances	Atlas landscape, prepared track, modest tents; white jellabas, turbans, long rifles, embroidered saddle cloths, decorated bridles; adults and children on both sidelines, men cheering, women ululating
	Classification	Odd-numbered Sorba of fifteen; Mqedem within the middle of the line, never ahead of it, marked by his rifle and his horse's ornaments; horsemen as peers, alike in clothing and posture
	Analytical	Attire, rifles, saddle cloths, and bridles "materially credible"; cultural attributes, not exotic decoration
	Symbolic	Mqedem's decorated rifle and horse ornaments as marks of leadership; raised rifles and smoke carrying the "inherited ceremonial meaning" of the event
Interactive	Contact	Mqedem's brief gaze toward the viewing position (restrained demand); horsemen concentrating on the shared action (offer); spectators looking at the riders
	Social distance	Whole line, horses, and rifles visible; neither a distant panorama nor a cropped portrait
	Perspective	Frontal, nearly eye-level viewpoint from beyond the track; moderately low camera height

Metafunction	Parameter	Instruction in the SPE prompt
	Modality	High naturalistic modality: believable daylight, ordinary shadows, consistent smoke, true-to-life proportions, restrained saturation, photographic detail; no cinematic effects
Compositional	Information value	Line as a continuous horizontal structure across the middle; center-margin, with the Mqedem as nucleus; rifles, smoke, and mountains above (Ideal), hooves, tack, track, and dust below (Real)
	Salience	Mqedem moderately salient through clearer facial detail, distinctive tack, and local contrast; never enlarged; line never broken
	Framing	Weak framing among riders (alignment, repeated white clothing, intersecting vectors); stronger separation between Sorba and spectators

3.4 Analytical Procedure

The semiotic analysis of the two images in Figure 1 (the first baseline and SPE renderings), checked against the other renderings in Figure 2, was conducted by the authors according to Kress and van Leeuwen's (2006) tripartite model, as summarized in the coding scheme presented in Appendix A. Each image was assessed for representational depth (how richly the image is composed narratively and conceptually), interactive meaning (how the image communicates a social relationship of the viewer to the image itself), and compositional organization (how coherent and purposeful all resources of informational value, salience or framing of the image are). Since the authors also designed the SPE framework, the analysis is open to confirmation bias; for this reason, Sections 4.2 to 4.4 report deviations from the prompts in both conditions, and the analysis should be read as an illustrative interpretation rather than as an independent evaluation.

4. Results and Discussion

4.1 Baseline Renderings vs SPE Renderings

Figure 1. Baseline and SPE Renderings of the Moroccan Tbourida (First Rendering of Each Condition)



Figure 1A. Baseline rendering



Figure 1B. SPE rendering

Note. Both images were generated by the authors using the Nano Banana image generator in the Gemini application (Google, 2026), and each is the first rendering produced by its prompt. Figure 1A = baseline prompt (18 October 2026); Figure 1B = SPE prompt (18 October 2026). Both prompts are reproduced in full in Appendix B; the other renderings are shown in Figure 2. The depicted event is the Moroccan Tbourida (gunpowder game), set in the Atlas Mountains.

Figure 2. The Other Renderings Obtained From the Two Prompts



Figure 2A. Baseline, rendering 2



Figure 2B. Baseline, rendering 3



Figure 2C. SPE, rendering 2



Figure 2D. SPE, rendering 3

Note. Renderings are numbered in the order in which they were generated (baseline: 18 October 2026; SPE: 18 October 2026), each in a fresh chat session; the first rendering of each condition is shown in Figure 1.

4.2 Comparative Analysis: Representational Processes

The representational metafunction serves as the "ideational" component of the image, specifying the participants, the actions they perform, and the circumstances surrounding them.

4.2.1 Narrative Processes: The Dynamics of the Charge

Narrative processes can be represented as vectors or force lines representing the unfolding of an action or event.
Concerning action processes, the vectors depicting the horsemen in the baseline prompt image are essentially diagonal, representing the horsemen moving across the field, with rifles raised at different angles and shots fired at different moments, and lacking a clear transactional structure. Conversely, the vectors in the semiotically prompt-engineered image are more complicated. The charge is directed straight at the viewer, and the rifles of all the horsemen form parallel vertical vectors at the moment of a simultaneous shot. The Mqedem at the center of the image is the "Actor" who is creating a vertical vector through his shouting and lifting of the rifle to achieve his "Goal" of a simultaneous shot of all the horsemen on either side of him. The same structure is found in the other two SPE renderings (Figure 2), and in none of the baseline renderings.
On the other hand, in reaction processes in the baseline prompt image, the horsemen are gazing ahead along the wrongly depicted line of the charge, while the reaction process is carried mainly by the large crowd on the left, whose members wave, clap, and cheer toward the riders. Although the SPE prompt specified that the horsemen should concentrate on the shared action rather than on the viewer, several horsemen in the SPE image seem to look toward the camera; the reaction process is again carried by the spectators at the margins, who are waving and cheering with their gazes directed toward the horsemen. In both images, the narrative of the spectators, who are focused collectively on the Tbourida's ritual importance, is indicated through these gazes and gestures.
Within the SPE framework, the Mqedem has been clearly identified as the "Speaker" in the mental/speech processes of this performance. In Figure 1B, the Mqedem is depicted shouting very loudly with his mouth open while prompting the horsemen to carry out their actions. In the baseline image, a shouting rider can also be seen, but nothing in the composition identifies him as the leader who gives the signal. The shout is also visible in the second SPE rendering, but not in the third, where the Mqedem's mouth is closed (Figure 2). This allows the "Mqedem" to function as a culturally grounded model of communicative leadership, offering lower-proficiency EFL learners a non-verbal entry point for meaning-making before verbal expression is required.

By including an audience at the margins, both images move the Tbourida from performance to communal event, prompting students to consider who is present, why, and what that presence signifies: a culturally grounded inference task with direct implications for intercultural competence.

4.2.2 Conceptual Processes: The Structure of Tradition

Conceptual processes describe representative roles assigned to the participant based on categories of class, structure, or meaning as opposed to what they do.

Images produced by the classification process in the SPE rendering provide a visual representation of their taxonomy, including ranks of Tbourida participants and their horses. As shown by the image, the Mqedem is rendered at the center of the line, between his horsemen, in a darker striped jellaba and a black cape and on the most richly ornamented horse, because he is the single-level superordinate; thus, the hierarchy is defined for Tbourida. The horsemen, depicted on either side of him, are wearing similar clothes and riding similarly ornamented horses, thus indicating that they are peers and subordinate to the Mqedem. In the other two SPE renderings, the Mqedem is again inside the line, set apart in one case by a brown cloak and in the other by a cream-colored jellaba and a grey horse with red tassels (Figure 2). In the baseline image, by contrast, the riders form one undifferentiated line, and no participant is set apart as superordinate, which is also the case in the other two baseline renderings. The size of the Sorba, however, departs from both prompts: the SPE renderings show twelve, fifteen, and fourteen riders, so that only the second one follows the instruction of an odd number of fifteen, and the baseline renderings show about ten riders each, their exact number being hard to establish where the horses overlap.

The analytical process within SPE images represents a relationship between a part and a whole; therefore, the horsemen and Mqedem are defined as "carriers," and their elaborately decorated saddles, Jellabas, turbans, and rifles represent their "possessive attributes." In this respect, the two images are closer than one might expect: both render the jellabas, turbans, embroidered saddle cloths, and tasselled bridles in fine detail, since both prompts describe them, and in both the rendered possessive attributes may reflect the principle of representational sovereignty. The difference lies in how they are organized. In the SPE image, the Mqedem's attributes, namely his striped jellaba and black cape and the green, red, and gold breastplate of his horse, set him apart from his horsemen, whereas in the baseline image the attributes of all the riders carry the same weight.

Concerning symbolic processes, the semiotically engineered visual focuses on the symbolic attribute: the Mqedem's decorated rifle, raised at the center of the line by the one rider who is singled out, is shown as the mark of his leadership, and the raised rifles and the smoke above the line carry what the SPE prompt called the "inherited ceremonial meaning" of the event, through prominence and modality. This elevates what is rendered in the baseline prompt image, where it is one rifle among many, as a "picture of a gun" to a "representation of history" in the SPE rendering, and potentially supports a key pedagogical objective of enhancing intercultural representation and understanding.

4.3 Comparative Analysis: Interactive Processes

The interactive meaning regulates the social relationship established between the producer, the viewer, and the represented participants.

4.3.1 Contact (Gaze): Demand vs. Offer

The baseline rendering serves primarily as an "offer". In this image, the horsemen are depicted in three-quarter view, looking ahead along the charge, thus not aware of the viewer. This renders the culture more as an object of interest viewed from a distance and gives the image a "tokenistic" quality. On the other hand, the semiotically engineered rendering creates a strong demand gaze. The Mqedem is positioned at the center, riding toward the viewer and looking directly towards him or her, demanding recognition, reflection, and respect for the ancestral art of Tbourida. The viewer is then socially obligated to recognize the Moroccan horsemen's subjectivity and vectors because of the nature of the demand gaze. Several of the horsemen beside him, however, also face the viewer, which weakens the contrast that the prompt intended between the gaze of the leader and that of his riders. The demand is not equally strong in the other SPE renderings: in the second, the Mqedem shouts with his eyes narrowed, and in the third he looks at the viewer calmly, with his mouth closed (Figure 2).

4.3.2 Social Distance (Size of Frame)

The baseline image is taken in a medium-long shot that shows the whole charge together with the crowd. However, by placing the viewer at a distance in a public space, the image offers fewer cues for emotional connection between the performers and the audience.

The SPE prompt asked for enough social distance to see the whole line, and this is what the rendering shows: a medium-long shot in which the whole Sorba fills the width of the frame, with nothing between the viewer and the horses. The Mqedem is no closer to the viewer than his horsemen, since he rides inside the line; it is the frontal charge, coming straight at the camera, that shortens the distance between the viewer and the troupe. This difference draws the viewer's attention to the Mqedem's serious facial expression as he calls the signal, creating a greater opportunity for the viewer to feel more connected to the Tbourida's performance, and potentially allowing students to develop a strong emotional connection to this performance instead of just observing the action.

4.3.3 Perspective (Power Dynamics)

In the baseline rendering, the horsemen are depicted from an oblique horizontal angle in three-quarter view, charging diagonally across the frame. This angle positions the viewer as a relatively detached observer of the event rather than someone directly involved in it (Kress & van Leeuwen, 2006). The SPE rendering places the viewer in the path of the charge and invites involvement as the horsemen are depicted at eye level from a frontal angle while charging. The generated rendering does not show a clear low angle although the SPE prompt asked for a moderately low camera height; instead, what conveys the Mqedem's superiority and authority is his central position, his darker dress, and the raised rifle that rises into the upper part of the frame—features that relate to salience and information value rather than to perspective (see Section 4.4). Together, the depictions of the Mqedem and the horsemen provide educators with material to discuss the structure governing "Tbourida" with their students.

4.3.4 Modality (Truth Value)

In comparison, the baseline rendering also exhibits a high naturalistic modality; it reads as a real event depiction rather than a cinematic scene despite the strongly saturated colors, particularly in the clothes of the crowd, the striped tents, and the flags. The SPE framework requires high-naturalistic modalities. Regarding realism and the overall level of detail represented on clothing, rifles, and Moroccan ornaments worn by the Mqedem, the SPE rendering appears as if captured by a professional photographer, particularly in the Mqedem and the horses at the center of the line. However, a closer look at the charge itself shows that the naturalism of the baseline rendering does not hold. The horses are squeezed into one overlapping mass where several riders cannot be matched with their horses, and some riders seem to float above the heads of the neighboring horses. Besides, the far end of the line appears to come out of the line of spectators, which gives the impression that the troupe started its charge from among the audience and not from the end of the track. This kind of incoherence, which is common in AI-generated crowds, lowers the modality of the charge, that is, the part of the image that carries the event. The two baseline renderings generated later (Figure 2) produce the same irregularities.

As far as lighting is concerned, the SPE prompt asked for realistic daylight and ordinary shadows, and the rendering complies, with a soft, slightly hazy light and faint shadows under the horses. The baseline prompt requested a bright sunny day whilst the corresponding rendering does not quite achieve this because the sky is white and hazy; only the third baseline rendering can provide the clear blue sky which was specifically asked for in the prompt (Figure 2). With regard to saturation levels, the baseline rendering is clearly higher in saturation while the SPE rendering, as required by the prompt, uses a more muted color palette featuring colors such as whites, browns, reds, and yellows. Moreover, the scale of the SPE rendering remains natural overall. The Mqedem is no larger than his horsemen, as the prompt required, and the image is perceived as one photograph and not as a composite collage. In this respect, the naturalistic modality was not compromised by the compositional parameters of the prompt.

4.4 Comparative Analysis: Compositional Processes

The compositional meaning relates to the layout of the visual and the way in which information is prioritized and connected.

4.4.1 Information Value: The Ideal and the Real

Generally, the baseline rendering lacks an organization of visual information in an orderly fashion. It shows its charge as a diagonal line crossing the frame, and although the baseline prompt explicitly asked for the Mqedem to ride in the middle of the line, the rendering does not distinguish him visually from the other horsemen. As a result, students may be unable to judge the structure of the event.

Figure 1B was created with a semiotic method, with a vertical axis being the method of encoding meaning: the raised rifles and the smoke of the shot rise into the upper part of the visual, against the mountains, with the Mqedem's rifle at their center, while

the galloping horses, the dust, and the track fill its lower part. According to Kress and van Leeuwen (2006), the position of the visual's top represents "Ideal" or the essence, instruction, and promise. The Mqedem guarantees the order and synchronization of the charge; the "Real," the galloping horses and the dusty ground at the bottom, serves as a physical basis of the investment and action of Tbourida. Along the horizontal axis, the Mqedem is the Centre and the horsemen on either side of him are the Margins, which is how the SPE prompt defined the information value of the line; the same organization is found in the other two SPE renderings (Figure 2).

4.4.2 Salience: Directing the Educational Gaze

In the baseline image, salience falls on the riders nearest to the camera, at the left end of the diagonal line, simply because of where they are; the horsemen are thus perceived as having equal importance, and none of them is singled out by his role. Thus, there is a loss of complexity within the cultural hierarchy and the use of illustrative convenience.

The salience of the semiotically-engineered imagery has been manipulated through placement, dress, and contrast. For example, the Mqedem's central position in the line, his darker striped jellaba and black cape against the white of his horsemen, and the ornaments of his horse make him the first figure the eye lands on, without making him larger than the others, which is exactly the moderate salience the prompt asked for. This structure is a message to the student that he is the single-level superordinate of the taxonomy, and is the basis for prompting the student to think about the significance of leadership in this ancestral art form.

4.4.3 Framing: Connection vs. Disconnection

In the baseline rendering, framing is mostly unintentional and inconsistent: the rope along the track separates the nearest riders from the crowd, yet the far end of the line merges with the spectators, and the dust partly separates the horsemen from the landscape. The semiotically engineered framework requested weak framing among the riders and a stronger separation between the Sorba and the spectators. In Figure 1B, the horsemen ride shoulder to shoulder, with the Mqedem among them, and nothing but the shared dust and smoke comes between them; the line, on the other hand, is set apart from the spectators by a strip of open ground and the ropes on both sides, which is precisely the contrast the prompt asked for.

4.5 Synthesis of Results

By combining a comparative analysis of representational, interactive, and compositional modes within one cultural case study, the study examines if semiotic prompt engineering (SPE) can positively reframe generative AI as much more than just a visually convenient tool used to illustrate ideas, as it exhibits pedagogical depth and advanced creation capabilities. Even though there is only one cultural case study and a handful of renderings per condition evaluated throughout the research, the qualitative differences noticed between the two renderings, whose prompts describe the same event, provide preliminary and theoretically driven support that the concepts behind SPE are plausible.

The baseline prompt generated a rich but undifferentiated image of a cultural event. Neither condition followed its prompt fully, however: the baseline renderings did not single out the Mqedem, although the prompt placed him in the middle of the line, they show fewer than the fifteen riders requested, their shots are not synchronized, and in the first rendering the riders are crowded into an incoherent, overlapping mass rather than lined up side by side, as the prompt described; the SPE renderings, for their part, show twelve and fourteen riders in two cases instead of an odd number of at least fifteen, do not show the moderately low camera height that was asked for, and in the third rendering the Mqedem does not shout. In contrast, SPE prompts create a visual representation by providing viewers with actions and reactions to consider, creating social distance for viewers based on how they perceive the actions, and creating an organization of information that reflects some of the values and hierarchy in the represented culture. Although the prompts describe the same event, the SPE prompt is longer and more prescriptive. The observed differences may reflect its semiotic instructions, additional detail, or organizational structure; their individual contributions cannot be isolated in this proof of concept (see Section 5.1). These findings should be considered solely as a proof of concept. Although the observed increases in narrative depth, social distance, and compositional hierarchy are consistent with Kress and van Leeuwen (2006), they must be verified independently to provide support for how generalized SPE may be as a pedagogical methodology via inter-rater coding, learner response studies, replications across multiple AI models, and the analysis of multiple generations per prompt.

It is worth noting that the semiotic organization of the SPE renderings was not achieved at the expense of ethnographic accuracy. In a real Tbourida charge, the Mqedem rides inside the line of horsemen, usually in its center, from where he gives the signal (UNESCO, 2021). Both prompts asked for this, but only the SPE renderings deliver it: in all three of them, the Mqedem is in the line, between his horsemen, whereas in the baseline renderings no leader can be found at all. This raises a point about

representational sovereignty: a visual can encode the values of a culture clearly while misrepresenting its practice, and the instruction about the Mqedem's place in the line seems to have been respected only when it was embedded in the compositional structure of the SPE prompt. Educators using SPE should therefore check the generated visuals against the cultural reality they depict, and may need to inform learners when a composition is symbolic rather than documentary.

Ultimately, the significance of these results will be attributed to their potential ability to provide educators with pedagogical value that should be integrated into creating culturally accurate visual depictions. Besides, they may provide active learning opportunities for learners to develop knowledge and understanding of culturally embedded meanings, and can be used to connect with, critique, and offer insights into these cultural representations. While baseline images of the Tbourida may have students thinking about the depicted performance, SPE images may have students thinking about the nature of the dynamic, leadership, community, heritage, and representation. The difference between these two examples suggests the distinction between convenient illustration of ideas and education through visual learning.

4.6 Pedagogical Implications

The differences described above have direct implications for the EFL classroom. First, the SPE image can be read as a visual narrative: students can be guided to trace the Mqedem's gesture to the horsemen on either side of him, practicing the skill of reading causality and sequence in images, and they can be asked whom the participants are looking at, and why, which opens a discussion of hierarchy, roles, and collective action.

Second, learners are given a chance to deduce the Mqedem's leadership role through the visual taxonomy. The precisely rendered attributes also allow teachers and students to treat the image as a cultural inventory, naming the jellaba, the turban, or the ornamented saddle, thus developing content-specific vocabulary instead of surface impressions. Also, elevating the rifle from object to symbol gives teachers a concrete way of introducing cultural symbolism, since the same object carries different meanings in different cultural contexts.

Third, the interactive choices may change how learners position themselves. Through the shift from offer to demand, students may move from being distant observers to being active participants in cultural meaning-making, which intercultural education frameworks consider central to the development of cross-cultural empathy (Byram, 1997). In addition, the frontal charge, which brings the whole troupe toward the learner, may create a stronger connection, since students tend to invest more in content that they find personally relevant (Fredricks et al., 2004). Naturalistic modality is also important in this regard, because an inclusive EFL environment aims at breaking down the orientalist gaze that cinematic imagery may reinforce (Said, 1978).

Finally, the compositional organization provides educators with an explicit analytical framework that can be shared with students. Asking students why certain components of the image are placed where they are may help them develop the critical visual literacy skills that contemporary curricula increasingly consider essential, alongside the traditional reading and writing competencies (New London Group, 1996; Serafini, 2014). Even before any verbal explanation, students who look at the image in sequence, from the leader to the collective, already start to learn how the Tbourida is organized.

5. Limitations and Recommendations

5.1 Limitations of the Study

As a proof of concept, this study demonstrates theoretically coherent and qualitatively observable differences between semiotically informed and baseline prompting. However, several important limitations constrain the scope of claims that can be drawn from these initial findings. First, even when applying semiotic constraints, generative AIs can produce hallucinations and inconsistencies or merge different cultural markers that need to be checked by humans. Second, operationalizing the SPE framework successfully requires an understanding not only of social semiotic theory, but also of the tenets of prompt engineering; many educators may face a steep learning curve in this area. Third, using Nano Banana through the Gemini application, without a verified record of the exact image-model variant, means that the generated results may not be replicable using other generative image models or later variants, due to differences in their underlying training data and their respective levels of response complexity for detailed compositional prompt instructions. Fourth, measures of quality, like narrative depth and cultural resonance, are subjective, and the present analysis was conducted by the authors who designed the SPE framework, which means that quantitative studies measuring student engagement with those aspects are needed to establish reliability and confidence in the data from this study. Fifth, generative AI models do not guarantee identical outputs from the same prompt across separate sessions, and only three renderings per prompt condition were generated in this study, which means that the specific images analyzed in this study represent a few instantiations of what the SPE framework can produce rather than a fixed and reproducible result; future studies should account for this variability by analyzing a larger number of generations per prompt condition. Sixth, the model may override explicit instructions in favor of default visual patterns, as in the absence of any identifiable Mqedem in the baseline renderings despite the instruction to place him in the middle of the line, or in the number of

riders, which fell short of the requested fifteen in five of the six renderings, which calls for the cultural verification of generated visuals by informed members of the represented community. Seventh, the 479-word baseline and 608-word SPE prompts differ in length, organization, and content: the SPE prompt is divided into three sections headed by the metafunctions and uses the terminology of the framework. The study therefore cannot separate the effects of additional detail, semiotic instructions, or the structure in which they are presented.

5.2 Recommendations for Educators and Curriculum Designers

The following recommendations can help educators use generative AI to curate educational materials. Rather than relying on trial-and-error natural language prompting, teachers may find the SPE process useful for ensuring that generated visuals reflect the representational, interactive, and compositional needs of their curricula. When creating localized content, like the Moroccan Tbourida, practitioners need to be specific about possessive attributes, like types of clothing, accessories, and environments, in order to avoid producing a generic representation of the content or an orientalist representation of the culture being depicted. Teacher training programs may now include semiotic prompt engineering as a requirement for EFL teacher trainees, enabling them to move from being passive curators to being informed creators of materials.

Finally, it is highly recommended that educational institutions create prompt repositories that contain validated, semiotically informed designs and provide teachers with access to those materials using a variety of cultural and scientific topics.

6. Conclusion

The findings of this research indicate that using Kress and van Leeuwen's (2006) social semiotic framework in conjunction with generative AI may improve the methodology of curating educational visuals. By focusing on how to create visuals using Semiotic Prompt Engineering (SPE) rather than illustrative convenience, educators have the ability to create instructional visuals that are not only aesthetic but also complex and useful for instructional purposes. The comparative analysis of Tbourida renderings suggests a greater depth of narrative, clearer social distance, and a stronger representation of societal hierarchy in visuals created with an understanding of representational, interactive, and compositional metafunctions than in those created via the shorter baseline prompt.

This project does not claim to have solved the methodological issues surrounding AI-assisted development of educational materials, but provides an initial theoretical basis for future research through a preliminary articulation of SPE as a framework, and encourages researchers in the field to put it to rigorous empirical testing, as the proposed application of this research demands. This project may ultimately contribute most to the field of education through the questions that it will generate: How does SPE perform in different cultural contexts? How will students respond to semiotic versus non-semiotic visuals? How can SPE be taught to educators with no background in social semiotics? These represent some of the specific areas for which this proof-of-concept has been designed to generate future academic research direction.

In conclusion, this research represents a new phase of generative curation that leverages the combined strength of human theoretical knowledge and AI's generative abilities to create inclusive, culturally appropriate, and pedagogically sound resources that better reflect student identities.

Statements and Declarations

Conflicts of Interest Statement

The authors declare no conflicts of interest in relation to the research, authorship, or publication of this article.

Ethics Statement

This study did not involve human participants.

Funding Statement

This research received no external funding.

Data Availability Statement

The prompts used in this study are reproduced in full in Appendix B; the generated images are shown in Figures 1 and 2 and are available in full resolution from the corresponding author upon reasonable request.

AI Use Disclosure Statement

Generative AI was used in this study. The images analyzed, which were the objects of study, were generated using the Nano Banana image generator through the Gemini application (Google, 2026; Google DeepMind, 2026) through structured semiotic prompts developed by the authors as part of their research design (Sections 3.1–3.3). The semiotic analysis was conducted by the authors, who verified all AI-assisted output and take full responsibility for the content of the article.

References

Abbas, N. F., Jasim, M. M., & Hussein, K. A. (2026). A social semiotic analysis of Instagram as a multimodal platform for supporting EFL learners. *Arab World English Journal*, *17*(1), 49–65. https://doi.org/10.24093/awej/vol17no1.4

Albeihi, H. H. M., & Rice, M. F. (2025). Generative AI and language diversity: Implications for teachers and learners. *Arab World English Journal*, *16*(1), 43–54. https://doi.org/10.24093/awej/vol16no1.3

Bianchi, F., Kalluri, P., Durmus, E., Ladhak, F., Cheng, M., Nozza, D., Hashimoto, T., Jurafsky, D., Zou, J., & Caliskan, A. (2023). Easily accessible text-to-image generation amplifies demographic stereotypes at large scale. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery. https://doi.org/10.1145/3593013.3594095

Bozkurt, A., Xiao, J., Farrow, R., Bai, J. Y. H., Nerantzi, C., Moore, S., Dron, J., Stracke, C. M., Singh, L., Crompton, H., Koutropoulos, A., Terentev, E., Pazurek, A., Nichols, M., Sidorkin, A. M., Costello, E., Watson, S., Mulligan, D., Honeychurch, S., . . . Asino, T. I. (2024). The manifesto for teaching and learning in a time of generative AI: A critical collective stance to better navigate the future. *Open Praxis*, *16*(4), 487–513. https://doi.org/10.55982/openpraxis.16.4.777

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., . . . Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.

Cazzaniga, L. (2026). *SCHEMA for Gemini 3 Pro Image: A structured methodology for controlled AI image generation on Google's native multimodal model* (arXiv:2602.18903). arXiv. https://doi.org/10.48550/arXiv.2602.18903

El Farouki, H., & Bouziane, K. (2025). *Is ChatGPT more Moroccan than me? Evaluating Semiotic Engine GPT's interpretation of Moroccan cultural artifacts*. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5174375

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Google. (2026). *Gemini (October 2026 version) [Generative AI application with image-generation features]*. https://gemini.google.com

Google DeepMind. (2026). Gemini Image: Nano Banana. https://deepmind.google/models/gemini-image/

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Edward Arnold.

Jewitt, C. (Ed.). (2009). *The Routledge handbook of multimodal analysis*. Routledge.

Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.

Liu, H., Liu, L., & Li, H. (2024). Multimodal discourse studies in the international academic community (1997–2023): A bibliometric analysis. *SAGE Open, 14*(4). https://doi.org/10.1177/21582440241305454

Liu, V., & Chilton, L. B. (2022). Design guidelines for prompt engineering text-to-image generative models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. https://doi.org/10.1145/3491102.3501825

Machin, D. (2007). *Introduction to multimodal analysis*. Hodder Arnold.

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66*(1), 60–92. https://doi.org/10.17763/haer.66.1.17370n67v22j160u

Ofosu-Asare, Y. (2025). Cognitive imperialism in artificial intelligence: Counteracting bias with indigenous epistemologies. *AI & Society,* 40(4), 3045–3061. https://doi.org/10.1007/s00146-024-02065-0

Oppenlaender, J. (2024). A taxonomy of prompt modifiers for text-to-image generation. *Behaviour & Information Technology, 43*(15), 3763–3776. https://doi.org/10.1080/0144929X.2023.2286532

Petty, S., & Benbouazza, B. (2020). Representational sovereignty in Moroccan Amazigh documentary film. *Studies in American Indian Literatures*, *32*(3–4), 129–148. https://doi.org/10.1353/ail.2020.0021

Qadri, R., Shelby, R., Bennett, C. L., & Denton, E. (2023). AI's regimes of representation: A community-centered study of text-to-image models in South Asia. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery. https://doi.org/10.1145/3593013.3594016

Qian, Y. (2025). Pedagogical applications of generative AI in higher education: A systematic review of the field. *TechTrends, 69*(5), 1105–1120. https://doi.org/10.1007/s11528-025-01100-1

Said, E. W. (1978). *Orientalism*. Pantheon Books.

Serafini, F. (2014). *Reading the visual: An introduction to teaching multimodal literacy*. Teachers College Press.

UNESCO. (2021). *Tbourida*. Intangible Cultural Heritage. https://ich.unesco.org/en/RL/tbourida-01483

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E. H., Le, Q. V., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, *35*, 24824–24837.

Weninger, C., & Kiss, T. (2013). Culture in English as a foreign language (EFL) textbooks: A semiotic approach. *TESOL Quarterly, 47*(4), 694–716. https://doi.org/10.1002/tesq.87

Appendix A

Table A1

Semiotic Coding Scheme and SPE Checklist

Metafunction	Category	Options (Kress & van Leeuwen, 2006)	Guiding question for prompt design
Representational	Action process	Transactional (Actor–vector–Goal); non-transactional (Actor–vector); event (vector–Goal)	Is there an Actor whose movement, gesture, or gaze forms a vector that tells the story of the event?
	Reaction process	Transactional (Reactor–gaze–Phenomenon); non-transactional (Reactor–gaze)	Do participants react to another participant or an off-frame event, and does their gaze guide the viewer?
	Mental and speech process	Senser (mental process); Speaker (speech process)	Is a participant shown speaking or commanding, giving a cause for the others' actions?
	Circumstances	Setting; means; accompaniment	Are the setting, the culturally specific objects, and secondary participants named and described?
	Classification process	Covert taxonomy; overt taxonomy (single-levelled, multi-levelled)	Is the hierarchy among participants visible, with participants of the same rank alike?
	Analytical process	Carrier (whole) and Possessive Attributes (parts)	Are culturally significant attributes (attire, tools, ornaments) rendered precisely?
	Symbolic process	Symbolic attributive; symbolic suggestive	Is an object carrying cultural or historical meaning made prominent?
Interactive	Contact	Demand; offer	Should a participant look at the viewer (demand) or away (offer), and why?
	Social distance	Close-up (intimate/personal); medium shot (social); long shot (impersonal)	Is the frame size suited to the closeness intended for the learner?
	Horizontal angle	Frontal (involvement); oblique (detachment)	Should the viewer be involved (frontal) or detached (oblique)?
	Vertical angle	High angle (viewer power); eye level (equality); low angle (represented participant power)	Is the vertical angle used deliberately to encode hierarchy?
	Modality	Detail; contextualization; depth; illumination; color	Are detail, setting, depth, light, and color naturalistic rather than cinematic or staged?
Compositional	Information value	Given/New (left/right); Ideal/Real (top/bottom); Center/Margin	Is placement (left/right, top/bottom, center/margin) intended rather than left to the AI's default?
	Salience	Size; placement; contrast; sharpness; color	Is the most important participant also the most salient, through placement, proximity, sharpness, contrast, or size?
	Framing	Strong (disconnection); weak (connection)	Are the connections and separations between elements intended?

Note. Adapted from Kress and van Leeuwen (2006). The guiding questions are meant to be answered in order, from the representational to the compositional metafunction, before the final prompt is composed.

Appendix B

Prompts Used in the Study

B1. Baseline Prompt

Create a horizontal 4:3 image of Tbourida that I can use as a teaching aid for my students in English classes. I want it to look like a real festival in Morocco, not just a picture of people on horses taken some time ago. The background should include an extended plain where the event takes place and where families watch. The sky should be bright and clear.

Tbourida is a great example of traditional Moroccan equestrian art also known as a "fantasia". This art has been passed on from generation to generation and was recognized as intangible cultural heritage by UNESCO in 2021. Tbourida can be seen during moussems, celebrations, weddings, and many other celebrations. For lots of families, being present at Tbourida is one of the most enjoyable activities during their family gatherings. You should also remember to make a Sorba show of horsemen galloping together. A Sorba traditionally consists of an odd number of riders (usually

15), who start riding together, moving along their designated track. At the end of the track, the horse riders shoot with blank shots from their guns. People observe how they shoot, and it is one of the reasons why Tbourida raises interest among the audience. People admire how well the horsemen are coordinated in their actions.

There is a leader of the show known as Mqedem, who coordinates the Sorba and gives the signal. The Mqedem rides with other horsemen in the middle of the Sorba rather than in front. His horse may have some decorative details, but the other horsemen do not really differ from Mqedem, as they are all wearing similar clothing. Their clothes have to be traditional and consist of white jellabas or draped clothes, white turbans, and other riding accessories.

Apart from that, you have to take into account the horses that matter greatly during Tbourida. Their community spends years to train them and months to make their saddles. It is necessary to make the horses look realistic and strong, as expected of horses, so they should have four legs positioned sensibly and not look like AI horses put together. The charge ought to be true to life and closer to the one used in Tbourida in Morocco.

Besides horses, the event involves people too. Thus, it is essential to represent the crowd along the sides of the prepared track. It can comprise men, women, and children, who enjoy the event, with some clapping and cheering while others are watching attentively or ululating. It can help to keep the crowd at a reasonable safety distance from the horses. Including some tents and other realistic details can make the scene more convincing and dynamic without turning it into a show.

Most importantly, the image should be reflective of the Moroccan tradition and carry some significance, respecting this culture and still being easy to comprehend by secondary school students..

B2. Semiotic Prompt Engineering (SPE) Prompt

Create a single photorealistic, documentary-style educational image of Moroccan Tbourida, in a 4:3 horizontal frame, depicting the Talqa/Qersa, which is the coordinated blank-rifle discharge during a Sorba's galloping charge, on an open plain near the Atlas Mountains. Treat this as a visual communication problem informed by Kress and van Leeuwen's three metafunctions, not as an opportunity to create a merely decorative spectacle. Preserve the historical and cultural specificity of a tradition recognized by UNESCO in 2021, practiced at moussems, celebrations, and community festivals. Depict an odd-numbered Sorba of around fifteen riders. The leader, or Mqedem, must ride within the middle of the same horizontal line; never position him several horse-lengths ahead.

Representational metafunction. Construct a coherent narrative process with the riders and horses functioning as Actors and the charge as a set of convergent forward movement vectors. Capture a realistic, single, legible instant immediately following the collective firing of blank cartridges: rifle barrels angled skyward, white realistic smoke emerging simultaneously, horses galloping with anatomically plausible strides, and dust recording the direction of travel. The shared action must be identifiable as coordination rather than isolated shooting. Render the Mqedem's commanding gesture and recently shouted signal as evidence of a speech process, but do not invent impossible simultaneous gestures. His decorated rifle and the distinctive ornamentation of his horse communicate leadership within the group, not separation from it. Represent the remaining horsemen as peers, coordinated in clothing and posture. Their white jellabas, turbans, traditional long rifles, embroidered saddle cloths, decorated bridles, and practical riding accessories must appear materially credible. These are recognizable cultural attributes, not interchangeable exotic decorations. Include spectators on both safe sidelines: adults and children attending together, cheering men, attentive observers, and women ululating. Their gestures and gazes establish reaction processes directed toward the Sorba. Show the Atlas landscape, prepared track, and festival tents as circumstances grounding the performance.

Interactive metafunction. Adopt a frontal, nearly eye-level viewpoint from a safely placed position beyond the track boundary, facing the advancing Sorba. Use a moderately low camera height so the riders project dignity without appearing monumental. Maintain sufficient social distance to see the entire line, horses, and rifles clearly; avoid both an impersonal distant panorama and an intimate portrait that crops the troupe. Give the Mqedem a brief, plausible gaze toward the viewer's position, creating a restrained demand relationship, while the other riders concentrate their gazes on the shared action and thus predominantly offer information to the viewer. Spectators look toward the riders, not the camera. Use high naturalistic modality: believable Moroccan daylight, ordinary shadows, physically realistic smoke, true-to-life proportions, restrained color saturation, credible depth, and photographic detail. No cinematic spotlights, fantasy costumes, or excessive dramatization.

Compositional metafunction. Organize the line as a realistically horizontal structure across the middle of the image, preserving the Mqedem's actual position among the riders. Establish center–margin information value: the Mqedem is the semantic nucleus, while the other riders remain clearly connected rather than subordinate figures receding behind him. Make him moderately salient through the slightly clearer facial detail, distinctive yet authentic tack, and local contrast, never by artificially enlarging him or breaking the line. Use the upper zone for raised rifles, smoke, mountains, and inherited ceremonial meaning (Ideal); use the lower zone for hooves, tack, track, and dust; ie, the material practice (Real). Maintain weak framing among riders through shared true-to-life alignment, white clothing, and intersecting movement vectors; use stronger spatial separation between the Sorba and spectators for safety and clarity. Keep the focal hierarchy intelligible without visually erasing any rider. The finished image should communicate coordinated action, collective identity, and cultural continuity while remaining suitable for close reading in a Moroccan EFL classroom.