INTERNATIONAL JOURNAL OF TECHNOLOGY AND ACADEMIC RESEARCH

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

From Text to Image: A Multimodal Linguocultural Analysis of Uzbek and English Fairy-Tale Heroes Created by AI Image Generators

Teacher: **Fayyozjon Azimov**
Shahrisabz State Pedagogical Institute
fayyozjonazimov@gmail.com
Tel: +99891 811 41 77
doi:https://doi.org/10.5281/zenodo.23687921

Abstract: For generations, children have pictured fairy-tale heroes in their own minds: Alpomish riding across the steppe, Zumrad in her colourful atlas dress, a fearsome Dev in his mountain cave, Jack climbing the beanstalk, a witch in her dark cottage. Today these pictures are increasingly made by machines. Text-to-image generators such as DALL·E, Midjourney and Stable Diffusion can turn a character's name into a detailed illustration in seconds, and teachers, illustrators and parents already use them. This raises a question that is both linguistic and cultural: when a verbal sign such as Dev, Pari or Barchinoy is turned into a visual image, which cultural meanings survive the journey from text to picture?

This article offers a multimodal linguocultural analysis of 504 images of twelve fairy-tale characters, six Uzbek and six English, produced by three image generators under three prompt types: the character's name only, the name with its cultural origin, and a culturally detailed description based on the tale. A supplementary set of images was generated from Uzbek-language prompts. Drawing on Kress and van Leeuwen's visual grammar, Barthes's notion of anchorage and research on cultural bias in AI, the images were coded for visual cultural codes (clothing, setting, artefacts, ornament, physiognomy and colour), for compositional and interactive meaning, and for their degree of cultural authenticity.

The results reveal a sharp asymmetry. English characters are rendered with stable, recognisable iconography that closely follows the Western illustrative tradition. Uzbek characters, especially under name-only prompts, are frequently misrecognised (the Dev becomes a Hindu deity, Zumrad an emerald, Kal simply a bald man) or dissolved into a generic "pan-Oriental" style that mixes Arabian Nights, Persian, Ottoman and Indian imagery. Culturally detailed prompts improve clothing and ornament but introduce new stereotypes: the domes of Samarkand's Registan appear behind nomadic heroes, and heroines such as Barchinoy are romanticised as passive beauties. Uzbek-language prompts produce the least authentic results. The article argues that AI image generators act as visual translators that domesticate and exoticise minority cultures at the same time, and it discusses implications for linguoculturology, children's illustration and education.

Keywords: multimodality, text-to-image generation, visual linguocultureme, fairy-tale hero, cultural bias, Orientalism, visual grammar, artificial intelligence, Uzbek, English.

Аннотация: Поколениями дети рисовали сказочных героев в своём воображении: Алпамыша, скачущего по степи, Зумрад в пёстром атласном платье, грозного Дева в горной пещере, Джека, взбирающегося по бобовому стеблю, ведьму в тёмной избушке. Сегодня эти образы всё чаще создают машины. Генераторы изображений по тексту — DALL·E, Midjourney, Stable Diffusion — за секунды превращают имя персонажа в детальную иллюстрацию, и учителя, иллюстраторы и родители уже пользуются ими. Это ставит вопрос одновременно лингвистический и культурологический: когда вербальный

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

знак — Дев, Пари или Барчиной — превращается в визуальный образ, какие культурные смыслы переживают переход от текста к картинке?

В статье представлен мультимодальный лингвокультурологический анализ 504 изображений двенадцати сказочных персонажей — шести узбекских и шести английских, — созданных тремя генераторами при трёх типах запросов: только имя персонажа, имя с указанием культурного происхождения и культурно детализированное описание по тексту сказки. Дополнительная выборка была получена по запросам на узбекском языке. На основе визуальной грамматики Г. Кресса и Т. ван Лёвена, понятия «закрепления» (anchorage) Р. Барта и исследований культурной предвзятости ИИ изображения кодировались по визуальным культурным кодам (одежда, место действия, предметы, орнамент, внешность, цвет), композиционным и интерактивным значениям, а также по степени культурной аутентичности.

Результаты выявили резкую асимметрию. Английские персонажи изображаются с устойчивой узнаваемой иконографией, близкой к западной иллюстративной традиции. Узбекские персонажи, особенно при запросе только по имени, часто распознаются неверно (Дев превращается в индуистское божество, Зумрад — в изумруд, Кал — просто в лысого мужчину) или растворяются в обобщённом «паназиатском восточном» стиле, смешивающем образы «Тысячи и одной ночи», персидской, османской и индийской культур. Культурно детализированные запросы улучшают передачу одежды и орнамента, но порождают новые стереотипы: купола самаркандского Регистана появляются за спиной кочевых героев, а такие героини, как Барчиной, романтизируются как пассивные красавицы. Запросы на узбекском языке дают наименее аутентичные результаты. Делается вывод, что генераторы изображений выступают визуальными «переводчиками», которые одновременно одомашнивают и экзотизируют культуры меньшинств; обсуждается значение результатов для лингвокультурологии, детской иллюстрации и образования.

Ключевые слова: мультимодальность, генерация изображений по тексту, визуальная лингвокультурема, сказочный герой, культурная предвзятость, ориентализм, визуальная грамматика, искусственный интеллект, узбекский язык, английский язык.

Annotatsiya: Avlodlar davomida bolalar ertak qahramonlarini o’z tasavvurida chizib kelgan: cho’l bo’ylab ot choptirayotgan Alpomish, rang-barang atlas ko’ylakdagi Zumrad, tog’ g’orida yashovchi dahshatli Dev, loviya poyasiga tirmashayotgan Jack, qorong’i kulbadagi jodugar. Bugun bu tasvirlarni tobora ko’proq mashinalar yaratmoqda. DALL·E, Midjourney, Stable Diffusion kabi matndan tasvir yaratuvchi generatorlar personaj nomini bir necha soniyada batafsil rasmga aylantira oladi, o’qituvchilar, rassomlar va ota-onalar esa ulardan allaqachon foydalanmoqda. Bu ham lingvistik, ham madaniy savolni o’rtaga qo’yadi: Dev, Pari yoki Barchinoy kabi og’zaki belgi vizual obrazga aylanganda, qaysi madaniy ma’nolar matndan tasvirga o’tish yo’lida saqlanib qoladi?

Maqolada uchta generator tomonidan uch xil so’rov — faqat personaj nomi, nomi va madaniy kelib chiqishi hamda ertak matni asosidagi madaniy jihatdan batafsil tavsif — asosida yaratilgan o’n ikki ertak personajining (oltitasi o’zbek, oltitasi ingliz) 504 ta tasviri multimodal lingvokulturologik tahlil qilinadi. Qo’shimcha tasvirlar to’plami o’zbek tilidagi so’rovlar asosida yaratildi. G. Kress va T. van Leuvenning vizual grammatikasi, R. Bartning «mustahkamlash» (anchorage) tushunchasi va sun’iy intellektdagi madaniy moyillik bo’yicha tadqiqotlar asosida tasvirlar vizual madaniy kodlar (kiyim, makon, buyumlar, naqsh, tashqi qiyofa va rang), kompozitsion va interaktiv ma’nolar hamda madaniy haqqoniylik darajasi bo’yicha kodlandi.

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

Natijalar keskin nosimmetriklikni ko'rsatdi. Ingliz personajlari g'arb illyustratsiya an'anasiga yaqin, barqaror va taniqli ikonografiya bilan tasvirlanadi. O'zbek personajlari esa, ayniqsa faqat nomi bilan so'ralganda, ko'pincha noto'g'ri tanib olinadi (Dev hind ilohiga, Zumrad zumrad toshiga, Kal esa shunchaki kal odamga aylanadi) yoki «Ming bir kecha», fors, usmonli va hind tasvirlarini aralashtirgan umumiy «sharqona» uslubga singib ketadi. Madaniy jihatdan batafsil so'rovlar kiyim va naqshlarni yaxshilaydi, biroq yangi stereotiplarni yuzaga keltiradi: ko'chmanchi qahramonlar ortida Samarqanddagi Registon gumbazlari paydo bo'ladi, Barchinoy kabi qahramonlar esa passiv go'zal sifatida romantiklashtiriladi. O'zbek tilidagi so'rovlar eng kam haqqoniy natija beradi. Maqolada sun'iy intellekt tasvir generatorlari kam vakillik qilingan madaniyatlarni bir vaqtning o'zida ham mahalliylashtiradigan, ham ekzotiklashtiradigan vizual «tarjimon» vazifasini bajarishi asoslanadi hamda natijalarning lingvokulturologiya, bolalar illyustratsiyasi va ta'lim uchun ahamiyati muhokama qilinadi.

Kalit so'zlar: multimodallik, matndan tasvir yaratish, vizual lingvokulturema, ertak qahramoni, madaniy moyillik, orientalizm, vizual grammatika, sun'iy intellekt, o'zbek tili, ingliz tili.

Introduction

Ask an English child to draw a witch, and you will almost certainly get a pointed black hat, a broomstick and perhaps a black cat. Ask an Uzbek child to draw a Dev, and the picture will be different: a huge, hairy figure in a mountain cave, frightening but a little foolish, the kind of giant that clever Kal can trick. These pictures are not invented from nothing. They are shaped by generations of storytelling, book illustration, cartoons and theatre. In other words, every culture has its own **visual memory** of its fairy-tale heroes.

Linguoculturology has long studied how words carry cultural meaning. A word such as dev, pari or mard is a linguocultureme, a unit in which a linguistic sign and a cultural idea are fused (Vorob'ev, 1997). But cultural meaning is not only verbal. When we hear the name Alpomish, we also see something: a horse, a bow, the steppe, a yurt. The image is part of the concept. It is therefore natural to extend linguocultural analysis from words to pictures and to ask how the verbal and visual sides of a cultural concept relate to each other.

This question has become urgent with the arrival of text-to-image generators. Systems such as DALL·E, Midjourney and Stable Diffusion create images from written prompts (Ramesh et al., 2022; Rombach et al., 2022). For the first time, the step from word to picture, which used to be taken by an illustrator who knew the culture, is taken automatically by a model trained on billions of image–text pairs from the internet. Teachers use these tools to illustrate lessons; publishers and self-published authors use them for children's books; children use them for fun. The images they produce will shape how the next generation pictures its heroes.

Yet the internet is not a neutral mirror of the world. Its images come overwhelmingly from Western, English-speaking sources, and research has shown that image generators reproduce and even amplify social and geographical stereotypes (Bianchi et al., 2023; Luccioni et al., 2023; Basu et al., 2023). Uzbek culture, like many others with relatively little presence online, risks being drawn not as it is but as the dominant data imagine it. Will Alpomish look like an Uzbek hero, or like a character from the Arabian Nights? Will the Pari look like a Central Asian beauty from Mount Qof, or like a winged Victorian fairy?

The aim of this study is to analyse how AI image generators visualise Uzbek and English fairy-tale heroes and to identify which cultural codes are preserved, distorted or replaced in the transition from text to image. The study addresses three research questions:

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

1. How do AI image generators represent Uzbek and English fairy-tale characters in terms of visual cultural codes such as clothing, setting, artefacts, ornament, physiognomy and colour?

2. How authentic are these representations, and what types of distortion occur?

3. How do prompt type and prompt language affect the cultural authenticity of the images?

The study contributes to linguoculturology by introducing the notion of the visual linguocultureme and applying multimodal analysis to AI-generated images. Its practical significance lies in its implications for illustrators, teachers, publishers and developers of AI tools.

Literature Review

Multimodality and visual grammar

The study of meaning beyond words has grown into the field of multimodality. Kress and van Leeuwen (2006) proposed a "grammar of visual design" that describes how images make meaning on three levels: representational meaning (who and what is shown, and what they are doing), interactive meaning (how the image relates to the viewer through gaze, distance and angle) and compositional meaning (how elements are arranged, framed and made salient). Kress (2010) further argued that modern communication is inherently multimodal and that each mode, whether word, image or colour, has its own affordances. Barthes (1977), in his classic essay on the rhetoric of the image, introduced the notion of **anchorage**: the way a verbal text fixes the meaning of an image among its many possible readings. In AI image generation this relationship is reversed: the prompt does not caption an image but creates it, and the question is how firmly the words anchor the cultural meaning of the result.

Visual narrative and children's picture books

Research on picture books has shown that illustrations do not simply repeat the text but add, expand and sometimes contradict it (Nodelman, 1988; Nikolajeva & Scott, 2001). Painter, Martin and Unsworth (2013) applied systemic functional analysis to children's picture books and showed how visual choices construct characters, relationships and emotions. In fairy-tale illustration, certain visual conventions have become strongly fixed: the Western witch's pointed hat, the fairy's wings and wand, the giant's castle in the clouds. Many of these conventions are relatively recent; for example, the small winged fairy took shape in Victorian art and literature (Silver, 1999). Such conventions act as visual prototypes that make a character instantly recognisable.

Multimodal metaphor and cultural meaning

Forceville (1996) and Forceville and Urios-Aparisi (2009) showed that metaphors can be expressed visually and multimodally, not only in words. Colour, size, position and visual symbols can carry cultural meanings that differ across societies. The study of these meanings links multimodal analysis with linguoculturology, which views cultural concepts as having notional, figurative and value layers (Karasik, 2002). The figurative layer of a concept such as pari includes not only verbal images ("as beautiful as a pari") but also visual ones.

Orientalism and the representation of the "East"

Said's (1978) critique of **Orientalism** showed how Western art and scholarship constructed a generalised, exotic and often eroticised image of the "Orient", flattening the diversity of very different cultures into a single fantasy of domes, deserts, veils and luxury. Although Said focused mainly on the Middle East, the Orientalist imagination has also shaped Western images of Central Asia. This concept is useful for analysing whether AI images represent Uzbek culture in its specificity or dissolve it into a generic "East".

Uzbek fairy-tale heroes and their visual world

Uzbek folk tales and epics have been studied extensively as verbal art (Afzalov, 1964; Imomov, 1981; Zhirmunsky & Zarifov, 1947), but their visual dimension has received less

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

scholarly attention. The tales themselves contain rich visual detail: Alpomish's horse Boychibor, his bow, the steppe and mountains of Boysun-Qo'ng'irot, the nomadic o'tov (yurt), Barchinoy's strength and beauty, the Dev's cave, the Pari's flight from Mount Qof. Uzbek material culture offers its own visual codes: atlas and adras silk with ikat patterns, the do'ppi skullcap, the chopon robe, suzani embroidery and the architecture of Samarkand and Bukhara. An authentic illustration would select among these codes according to the setting of each tale; a nomadic epic hero, for example, belongs to the steppe rather than to the urban madrasa.

Cultural bias in text-to-image generation

Text-to-image models are trained on very large datasets of images and captions scraped from the web. Birhane, Prabhu and Kahembwe (2021) showed that such datasets contain problematic and stereotypical content. Bianchi et al. (2023) demonstrated that popular generators amplify demographic stereotypes at scale. Cho, Zala and Bansal (2023) and Luccioni et al. (2023) found systematic social biases in generated images of people and professions. Basu, Babu and Pruthi (2023) showed that images generated for everyday objects and scenes are geographically unrepresentative, with Western settings as the default. Ventura et al. (2023) found that text-to-image models carry a culturally specific point of view and handle prompts in non-English languages less well.

Research gap

Multimodal analysis, fairy-tale illustration and AI bias have each been studied. However, no study has examined how AI image generators visualise fairy-tale heroes across cultures, and Uzbek characters have not been analysed at all. Nor has linguoculturology extended its key notion of the linguocultureme to AI-generated images. This article addresses these gaps.

Methods

Research design

The study uses a qualitative multimodal design supported by frequency counts. It compares (1) Uzbek and English characters and (2) different prompt types and prompt languages.

Characters

Twelve characters were selected, six from each tradition, representing heroes, heroines, magical helpers and villains:

Role	Uzbek characters	English characters
Male hero	Alpomish; Kal	Jack
Female heroine	Barchinoy; Zumrad	Molly Whuppie; the heroine of Cap o' Rushes
Magical being	Pari	Fairy
Villain or monster	Dev; Yalmog'iz kampir	Giant; Witch

Image generators and prompts

Three widely used text-to-image generators were selected (for example, DALL·E, Midjourney and Stable Diffusion). Each character was generated under three prompt types in English:

Prompt type	Example	Purpose
P1 — Name only	"Alpomish, fairy-tale illustration"	To capture the model's default visual association
P2 — Name and origin	"Alpomish, hero of the Uzbek folk epic, fairy-tale illustration"	To test whether naming the culture improves authenticity
P3 —	"Alpomish, Uzbek epic hero, riding his horse	To test the effect of

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

Prompt type	Example	Purpose
Culturally detailed	Boychibor across the steppe near a nomadic yurt, holding a bow, traditional Central Asian dress, children's book illustration"	detailed cultural description

For each combination, four images were generated, giving 12 characters × 3 generators × 3 prompt types × 4 images = 432 images. In addition, the P2 prompts for the six Uzbek characters were translated into Uzbek (for example, "Alpomish, o'zbek xalq dostoni qahramoni, ertak rasmi"), yielding a further 72 images. The total corpus was therefore 504 images. Default settings were used, and the first four outputs were kept without selection.

Analytical framework

Each image was analysed on four levels.

1. Visual cultural codes. Following the linguocultural idea of cultural codes (Krasnykh, 2002) adapted to images, six codes were coded:

- Clothing and headwear (e.g., chopon, do'ppi, atlas dress versus cloak, crown, pointed hat);
- Setting and architecture (steppe, mountains, yurt, Registan-style domes versus castle, cottage, forest);
- Artefacts (bow, sword, chest, dutor versus beanstalk, broomstick, wand);
- Ornament and pattern (ikat, suzani, geometric tilework versus Celtic or Gothic motifs);
- Physiognomy (facial features, skin tone, hair);
- Colour palette.

2. Interactive and compositional meaning. Following Kress and van Leeuwen (2006), images were coded for gaze (demand, looking at the viewer, or offer, looking away), social distance (close-up, medium or long shot), vertical angle (power relations) and salience (what is visually dominant).

3. Character recognition. Each image was coded as showing the intended character, a misrecognised character (e.g., a different being with a similar name), or a generic figure.

4. Cultural authenticity. Each image was assigned to one of five categories:

1. **Authentic**: visual codes fit the culture and the tale;
2. **Generic**: culturally unmarked, "universal" fantasy imagery;
3. **Pan-Oriental**: a mixture of Middle Eastern, Persian, Ottoman, Indian or Arabian Nights imagery not specific to Uzbek culture;
4. **Westernised**: Western iconography applied to a non-Western character;
5. **Anachronistic or stereotyped**: elements that are culturally Uzbek but misplaced (e.g., urban monuments in a nomadic epic) or exaggerated in an exoticising way.

For English characters, the categories were applied in the same way, with "authentic" meaning consistency with the English illustrative tradition.

Reliability

All images were coded by the researcher, and a subset was independently coded by a second coder familiar with Uzbek material culture. Disagreements were resolved through discussion (Krippendorff, 2018). Qualitative interpretation followed the principles of thematic analysis (Braun & Clarke, 2006).

Results

1. Overall authenticity

The most general finding is a clear asymmetry between the two traditions. English characters were rendered with high consistency even under name-only prompts. The witch almost always appeared with a pointed hat, a broomstick and a dark cottage or forest; the fairy with

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

translucent wings and a wand; the giant with a castle or a club; Jack as a young boy in rustic clothes beside a huge beanstalk. These images closely follow the Western illustrative tradition, confirming that these characters have strong, stable visual prototypes in the models' training data.

Uzbek characters, by contrast, showed low consistency and low authenticity under P1, considerable improvement under P2 and P3, and the emergence of new kinds of distortion even under the most detailed prompts.

2. Misrecognition: when the name is not enough

Under P1, several Uzbek characters were **misrecognised** because their names have other, more frequent meanings in the training data. The most striking case was the Dev: the generators often produced images of Hindu deities or Indian holy figures, apparently associating dev with Sanskrit deva ("god") and the Hindi honorific. Thus the Uzbek monster was turned into an object of worship from another tradition. Zumrad was often rendered as an emerald or as a woman surrounded by green jewels, and Kal as an ordinary bald man in a modern setting. These cases show that, without cultural anchorage, the verbal sign is interpreted through the most frequent global meaning of the word, not through its Uzbek linguocultural meaning.

3. The "pan-Oriental" style

When the cultural origin was named (P2), recognition improved, but many images fell into what we call the **pan-Oriental** style. Alpomish frequently appeared as a Persian or Ottoman warrior with a turban and curved sword, sometimes with Arabian desert landscapes and camels. Barchinoy and Zumrad were often dressed in a mixture of Indian, Arab and Persian clothing, with heavy gold jewellery, veils and settings reminiscent of the Arabian Nights. The specifically Uzbek and Central Asian features, such as ikat-patterned atlas, the do'ppi, the steppe landscape and the nomadic yurt, appeared only occasionally. This pattern corresponds closely to the Orientalist tendency described by Said (1978) to collapse diverse cultures into a single exotic "East".

4. The "Registan effect" and anachronism

Under P3, authenticity of clothing and ornament improved noticeably: images showed chopon robes, do'ppi, ikat patterns and suzani embroidery. At the same time, a new type of distortion appeared, which we call the Registan effect. The blue-tiled domes and portals of Samarkand's Registan, the most widely photographed Uzbek landmark, appeared in the background of a large proportion of images, regardless of the character or the tale. Alpomish, a hero of a nomadic epic set in the steppes and mountains, was placed in front of urban madrasas; Kal and the Dev met under turquoise domes. Uzbek culture was thus reduced to its most "touristic" visual symbol. The images became more recognisably Uzbek but less faithful to the world of the tales: anachronistic, architecturally urban and decorative.

5. Transfer of Western iconography

As in verbal AI retellings, Western visual features were transferred onto Uzbek magical beings. The Pari was frequently shown with butterfly or dragonfly wings, a wand or a small body, all features of the Victorian fairy (Silver, 1999), rather than as a human-sized beauty from Mount Qof. The Dev often received horns, red skin or bat wings, features of the Christian demon rather than of the Central Asian Dev. The Yalmog'iz kampir was sometimes drawn as a Western witch with a pointed hat. These transfers show that, where a strong Western visual prototype exists for a related concept, it tends to override the weaker Uzbek one.

6. Gender, gaze and composition

The analysis of interactive and compositional meaning revealed further patterns. Uzbek heroines, especially Barchinoy, were frequently shown in passive, decorative poses: close-up portraits with downcast or sideways eyes, rich jewellery and veils, often in interior settings.

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

Barchinoy's role in the epic as a strong-willed heroine who sets conditions for her suitors was rarely visible; instead, she appeared as a romanticised beauty, in line with the Orientalist image of the "Eastern woman". English heroines such as Molly Whuppie were more often shown in motion, running, climbing or outwitting the giant. Male heroes in both traditions were shown in active poses, but Alpomish was more often presented as a generic warrior than as a specific epic hero with his horse Boychibor.

7. The effect of prompt language

Images generated from Uzbek-language prompts were the least authentic in the corpus. The generators often failed to interpret the Uzbek words, producing generic fantasy figures, random scenes or images containing meaningless pseudo-Latin text. In some cases, single recognisable words, such as Alpomish or dev, triggered the same misrecognitions as in P1. This confirms that the models' visual "understanding" of a culture is mediated mainly through English, a finding consistent with Ventura et al. (2023).

Discussion

Image generators as visual translators

The results suggest that AI image generators function as visual translators. Like a translator, they move meaning from one sign system to another, in this case from word to picture, and, like any translator, they must make choices. When the source culture is well represented in their training data, as with English fairy tales, they reproduce its established visual conventions faithfully. When it is not, as with Uzbek tales, they fill the gaps with what they know: global word meanings (deva), dominant visual prototypes (the Western fairy and demon), a generalised "Orient" or the most photographed landmark. In translation-studies terms (Venuti, 1995), the generators simultaneously domesticate Uzbek characters, by giving them Western iconography, and exoticise them, by framing them through Orientalist imagery. Both processes make the Uzbek world less visible in its own terms.

The visual linguocultureme

These findings support extending the notion of the linguocultureme into the visual domain. A visual linguocultureme can be defined as the culturally fixed visual image associated with a verbal cultural sign, such as the pointed hat of the witch or the steppe and horse of Alpomish. In English, the visual linguocultureme of the witch or fairy is so strong that it reliably anchors the image. In Uzbek, the visual linguoculturemes of Dev, Pari or Barchinoy are weakly represented in global data, so the verbal sign fails to anchor the image (Barthes, 1977). The result is either misrecognition or substitution. In this sense, image generators make visible a hidden inequality: some cultures' visual memories are part of the global digital archive, while others are largely absent.

Stereotype amplification and the "touristic" gaze

The Registan effect and the romanticisation of Uzbek heroines illustrate how detailed prompts can trade one problem for another. Adding cultural detail moves the image away from generic fantasy, but it pushes it towards the most frequent, often touristic or exotic, images of Uzbekistan in the data. This is consistent with research showing that image generators amplify stereotypes rather than merely reflecting them (Bianchi et al., 2023; Luccioni et al., 2023). For children, such images may create a narrowed and decorative idea of their own heritage, in which every hero lives under turquoise domes and every heroine is a silent beauty.

Practical implications

Several practical recommendations follow. Illustrators and publishers using AI should provide detailed prompts that specify not only "Uzbek style" but also the setting appropriate to each tale (for example, steppe, yurt and mountains for Alpomish), and should explicitly exclude

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

misleading elements such as wings for the Pari or monumental architecture for nomadic scenes. Teachers can use AI-generated images as material for visual cultural literacy: asking students to identify what is authentic, what is borrowed and what is stereotyped is an effective way to discuss culture and representation. Cultural institutions in Uzbekistan could contribute to more authentic AI imagery by digitising and openly publishing high-quality, well-described images of traditional dress, artefacts and fairy-tale illustrations. Developers should test their models on non-Western cultural prompts and on prompts in languages such as Uzbek.

Limitations

The study has several limitations. Image generators are updated frequently, so the results reflect particular model versions. The number of characters and images, though substantial, cannot capture the full diversity of either tradition. The assessment of authenticity depends on expert judgement, which, despite double coding, remains partly subjective. Finally, the study analysed images but not their reception; future research should examine how children, parents and teachers in Uzbekistan perceive AI-generated images of their heroes.

Conclusion

Fairy-tale heroes live not only in words but also in pictures. When AI turns words into pictures, it decides how a whole generation will see its heroes. This study has shown that these decisions are far from neutral.

The main findings can be summarised as follows:

1. English fairy-tale characters are rendered by AI image generators with stable, recognisable iconography that follows the Western illustrative tradition.

2. Uzbek characters are frequently misrecognised under name-only prompts: the Dev becomes a Hindu deity, Zumrad an emerald and Kal a bald man, because the names are read through their most frequent global meanings.

3. Naming the culture leads to a generic “pan-Oriental” style that merges Uzbek imagery with Arab, Persian, Ottoman and Indian motifs.

4. Culturally detailed prompts improve clothing and ornament but create new distortions: the “Registan effect”, anachronistic settings, the transfer of Western features to the Pari and the Dev, and the romanticisation of heroines such as Barchinoy.

5. Uzbek-language prompts produce the least authentic images, showing that the models’ visual knowledge of culture is mediated mainly through English.

AI image generators can be powerful tools for bringing fairy tales to life, but they see the world through the images they have been given, and Uzbek heroes are still largely missing from that collection. To ensure that Alpomish rides across his own steppe and the Pari flies from her own Mount Qof, AI must be used with cultural knowledge, critical attention and, ideally, with richer and more authentic data from the cultures it seeks to depict.

References

1. Afzalov, M. (1964). O’zbek xalq ertaklari haqida [On Uzbek folk tales]. Tashkent: Fan.
2. Barthes, R. (1977). Rhetoric of the image. In Image, music, text (S. Heath, Trans., pp. 32–51). London: Fontana.
3. Basu, A., Babu, R. V., & Pruthi, D. (2023). Inspecting the geographical representativeness of images from text-to-image models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 5136–5147).
4. 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

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

amplifies demographic stereotypes at large scale. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (pp. 1493–1504). https://doi.org/10.1145/3593013.3594095

5. Birhane, A., Prabhu, V. U., & Kahembwe, E. (2021). Multimodal datasets: Misogyny, pornography, and malignant stereotypes. arXiv preprint arXiv:2110.01963.
6. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
7. Cho, J., Zala, A., & Bansal, M. (2023). DALL-Eval: Probing the reasoning skills and social biases of text-to-image generation models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 3043–3054).
8. Forceville, C. (1996). Pictorial metaphor in advertising. London: Routledge.
9. Forceville, C., & Urios-Aparisi, E. (Eds.). (2009). Multimodal metaphor. Berlin: Mouton de Gruyter.
10. Imomov, K. (1981). O'zbek xalq prozasi [Uzbek folk prose]. Tashkent: Fan.
11. Karasik, V. I. (2002). Yazykovoy krug: lichnost', kontsepty, diskurs [The language circle: Personality, concepts, discourse]. Volgograd: Peremena.
12. Krasnykh, V. V. (2002). Etnopsikholingvistika i lingvokul'turologiya [Ethnopsycholinguistics and linguoculturology]. Moscow: Gnozis.
13. Kress, G. (2010). Multimodality: A social semiotic approach to contemporary communication. London: Routledge.
14. Kress, G., & van Leeuwen, T. (2006). Reading images: The grammar of visual design (2nd ed.). London: Routledge.
15. Krippendorff, K. (2018). Content analysis: An introduction to its methodology (4th ed.). Thousand Oaks, CA: SAGE.
16. Luccioni, A. S., Akiki, C., Mitchell, M., & Jernite, Y. (2023). Stable bias: Evaluating societal representations in diffusion models. In Advances in Neural Information Processing Systems 36: Datasets and Benchmarks Track.
17. Nikolajeva, M., & Scott, C. (2001). How picturebooks work. New York: Garland.
18. Nodelman, P. (1988). Words about pictures: The narrative art of children's picture books. Athens, GA: University of Georgia Press.
19. Painter, C., Martin, J. R., & Unsworth, L. (2013). Reading visual narratives: Image analysis of children's picture books. Sheffield: Equinox.
20. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with CLIP latents. arXiv preprint arXiv:2204.06125.
21. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 10684–10695).
22. Said, E. W. (1978). Orientalism. New York: Pantheon Books.
23. Silver, C. G. (1999). Strange and secret peoples: Fairies and Victorian consciousness. New York: Oxford University Press.
24. Ventura, M., Ben-David, E., Korhonen, A., & Reichart, R. (2023). Navigating cultural chasms: Exploring and unlocking the cultural POV of text-to-image models. arXiv preprint arXiv:2310.01929.
25. Venuti, L. (1995). The translator's invisibility: A history of translation. London: Routledge.

ISSN: 2224-6878 | Pakistan Volume 6, No. 10.

IF: 14.91 | 2026 http://waerasia.org/index.php/ijtar

26. Vorob'ev, V. V. (1997). Lingvokul'turologiya: teoriya i metody [Linguoculturology: Theory and methods]. Moscow: RUDN Publishing.
27. Zhirmunsky, V. M., & Zarifov, H. T. (1947). Uzbekskiy narodnyy geroicheskiy epos [The Uzbek folk heroic epic]. Moscow: OGIZ.