DEVICE AND METHOD FOR GENERATING INDIVIDUALIZED IMAGES

TECHNICAL FIELD

[0001] Various aspects of this disclosure relate to devices and methods for generating individualized images.

BACKGROUND

[0002] Images, possibly annotated with text, have various applications like illustration, advertisements, warnings etc. Conventional approaches for the creation of visual assets, notably images, are arduous, involving a process several steps like organizing photo shoots, post-production photo edits, and engaging copywriters for taglines. Furthermore, depending on the use case (e.g. for advertisements), the generation of personalized (or individualized) images is desirable, but the intricate process of conventional approaches only yields a single, generalized image. To personalize it would necessitate repeating the above steps multiple times, exacerbating the resource and time constraints. When it comes to illustrations, the process typically begins with a design brief, followed by extensive research, days of drawing, and finalizing the artwork, all of which is time-consuming.

[0003] Accordingly, approaches for efficient generation of personalized images (in particular images annotated with text) are desirable.

SUMMARY

[0004] Various embodiments concern a method for generating individualized images, comprising training a text-to-image model, generating, by means of a large language model, a text prompt for the text-to-image model by instructing the large language model to provide a text prompt for the text-to-image model for a given context including at least one of a use case and a type of user for which the individualized image is to be generated and generating an individualized image for the given context by feeding the generated text prompt to the text-to-image model.

[0005] According to one embodiment, the type of user is a user persona.

[0006] According to one embodiment, determining the context comprises determining the user persona from a set of (different) user personas.

[0007] According to one embodiment, the method comprises generating the set of user personas by segmenting a set of given users.

[0008] According to one embodiment, the method comprises segmenting the set of given users by a machine learning clustering model.

[0009] According to one embodiment, the text-to-image model is a low rank adaption model of a stable diffusion model.

[0010] According to one embodiment, the method comprises generating the low rank adaption model from a predetermined stable diffusion model.

[0011] According to one embodiment, wherein generating the low rank adaption model comprises adapting the stable diffusion model to a predetermined image style.

[0012] According to one embodiment, the method comprises generating multiple low rank adaption models, wherein the low rank adaption models are associated with different styles and the individualized image is generated by feeding the generated text prompt to the text-to-image model associated with a desired image style.

[0013] According to one embodiment, the method comprises receiving a specification of the desired image style.

[0014] According to one embodiment, the method comprises receiving a specification of the context.

[0015] According to one embodiment, the method comprises generating a text annotation for the individualized image and including the text annotation into the individualized image.

[0016] According to one embodiment, the method comprises generating the text annotation depending on the given context.

[0017] According to one embodiment, the method comprises generating the text annotation by instructing the large language model to provide the text annotation for the given context.

[0018] According to one embodiment, the method further comprises stitching the individualized image to a predetermined background image.

[0019] According to one embodiment, a server computer is provided comprising a radio interface, a memory interface and a processing unit configured to perform the method of any one of the embodiments described above.

[0020] According to one embodiment, a computer program element is provided comprising program instructions, which, when executed by one or more processors, cause the

one or more processors to perform the method of any one of the embodiments described above.

[0021] According to one embodiment, a computer-readable medium is provided comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of the embodiments described above.

[0022] According to one embodiment, a computer program element is provided comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of the embodiments described above.

[0023] According to one embodiment, a computer-readable medium is provided comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of the embodiments described above.

[0024] It should be noted that embodiments described in context of an image generation system are analogously valid for the method for generating an individualized image and vice versa.

BRIEF DESCRIPTION OF THE DRAWINGS

[0025] The invention will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the accompanying drawings, in which:

- FIG. 1 shows a communication arrangement of a marketplace system.
- FIG. 2 illustrates an image personalization system according to an embodiment.
- FIG. 3 shows a flow diagram illustrating a method for generating individualized images.
- FIG. 4 shows a server computer according to an embodiment.

DETAILED DESCRIPTION

[0026] The following detailed description refers to the accompanying drawings that show, by way of illustration, specific details and embodiments in which the disclosure may be practiced. These embodiments are described in sufficient detail to enable those skilled in the

art to practice the disclosure. Other embodiments may be utilized and structural, and logical changes may be made without departing from the scope of the disclosure. The various embodiments are not necessarily mutually exclusive, as some embodiments can be combined with one or more other embodiments to form new embodiments.

[0027] Embodiments described in the context of one of the devices or methods are analogously valid for the other devices or methods. Similarly, embodiments described in the context of a device are analogously valid for a vehicle or a method, and vice-versa.

[0028] Features that are described in the context of an embodiment may correspondingly be applicable to the same or similar features in the other embodiments. Features that are described in the context of an embodiment may correspondingly be applicable to the other embodiments, even if not explicitly described in these other embodiments. Furthermore, additions and/or combinations and/or alternatives as described for a feature in the context of an embodiment may correspondingly be applicable to the same or similar feature in the other embodiments.

[0029] In the context of various embodiments, the articles “a”, “an” and “the” as used with regard to a feature or element include a reference to one or more of the features or elements.

[0030] As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

[0031] In the following, embodiments will be described in detail.

[0032] FIG. 1 shows a communication arrangement of a marketplace system, including a smartphone 100 and a server (computer) 106.

[0033] The smartphone 100 has a screen showing the graphical user interface (GUI) of an app for using one or more of various services, such as ordering food, which the smartphone’s user has previously installed on his smartphone and has opened (i.e. started) to use the service, e.g. to order food.

[0034] The GUI 101 includes graphical user interface elements 102, 103 helping the user to use the service, e.g. a map of a vicinity of the user’s position, food available in the user’s vicinity (which the app may determine based on a location service, e.g. a GPS-based location service), a button for placing an order, etc.

[0035] When the user has made a selection for a service, e.g. a selection of a restaurant or an online supermarket and/or a selection of food or groceries to order, the app communicates

with a server 106 of the respective service via a radio connection. The server may be seen as a part of a system implementing a service provider platform, i.e. a platform allowing users to access one or more services (food delivery, grocery delivery, e-hailing, etc.) The server 106 (carrying out a corresponding server program by means of a processor 107) may consult a memory 105 or a data storage 108 having information regarding the service (e.g. prices, availability, estimated time for delivery etc.) The server communicates any data relevant or requested by the user (such as estimated time for delivery) back to the smartphone 100 and the smartphone 100 displays this information on the GUI 101. The user may finally accept a service, e.g. order food. In that case, the server 106 informs the service provider 104, e.g. a restaurant or online supermarket accordingly. The server 106 may also communicate earlier with the service provider 104, e.g. for determining the estimated time for delivery.

[0036] It should be noted while the server 106 is described as a single server, its functionality, e.g. for providing a food delivery service will in practical application typically be provided by an arrangement (i.e. a server computer system) of multiple server computers (e.g. implementing a cloud service). Accordingly, the functionality described in the following provided by a server (e.g. server 106) may be understood to be provided by an arrangement of servers or server computers.

[0037] To make the service known to users, advertisements are often used. This may for example happen on the GUI 101, typically in a certain app, e.g. a dedicated app for accessing the service provider platform but possibly also other apps like a web browser, social media client app, email client app etc. For example, a banner 109 is displayed on the GUI 101 when the app is used.

[0038] Images play pivotal roles for advertisements ranging from in-app advertising banners, graphical tiles, icons, to external promotional channels like emails and social media. According to various embodiments, the efficiency and caliber of the production of images is elevated. When used for advertisements, the competitive edge of a service provider (e.g. food or grocery provider) can be fortified and superior user experience can be achieved.

[0039] According to various embodiments, to generate images (e.g. for advertisements like a banner 109) a LoRA (Low Rank Adaptation) is integrated with a SD (Stable Diffusion) model. Together, these tools (models) enable the generation of images enhanced with machine learning, e.g. to segment a user base (e.g. of the respective service for which advertisement images should be generated), and linear optimization techniques to refine

business use case prioritization, enhancing the content experience, e.g. within an app which allows accessing one or more services (e.g. delivery of groceries as well as food as well as e-hailing).

[0040] According to various embodiments, by delving deeper into user data, encompassing transactional histories and experiential insights, distinct behavioral patterns and preferences can be discerned. Such insights allow segmenting users into specific cohorts. With the assistance of an LLM (large language model), (image) content as well as suggestions can be generated meticulously suited to individual user needs, making their engagements with the respective service provider more personalized, be it through in-app promotional banners or social media marketing initiatives.

[0041] The potential for image generation is especially evident in the advertising and marketing sector, specifically when integrated with machine learning. Automated personalization, driven by an LLM, possesses broad adaptability and can span various applications. It is readily adaptable to customize content within an app which allows accessing one or more services (e.g. aimed at drivers and merchants). Moreover, various embodiments empower users in mobility, deliveries, financial services, advertisements and more to craft finely-tailored content. This is achieved by combining image generation capabilities with a separate personalized text generation service. In other words, according to various embodiments, there is an amalgamation of machine learning models, ranking algorithms, and LLMs for generating (in particular image) content.

[0042] A conventional one-size-fits-all approach to generate visual content leads to that, irrespective of their distinct preferences or personas, all users are served the same visual content. Thus, the visual content may align with the preferences of some users but could miss the mark for many others. For example, using a static image for an advertisement banner in a service provider app (which is universal for all users), while the user base is diverse with varied interests and personas, the visual content remains static and singular. This results in a mismatch where the content resonates with only a segment of the audience, leaving others with a possibly irrelevant or less engaging experience.

[0043] The approach for image generation according to an embodiment generates, instead of a universal (e.g. generalized) banner, personalized banners tailored to individual users. For example, for a discount campaign an embodiment would generate persona-specific banners (e.g. key visuals). A user with a family-oriented persona might see an image of a family

engaging with a service that is provided, while a gastronome gets visuals centered on food, and a regular e-hailing user would be greeted with transportation-centric imagery. This tailored approach ensures content relevance, enhancing user engagement and experience.

[0044] So, according to various embodiments, an image personalization system crafted for generation of image data (e.g. for a service provider to generate personalized visual advertisements) is provided. The system identifies and understands the distinct personas of users (e.g. users of the service), thereby enabling the customization of both imagery and textual content. Unlike, for example, displaying a standardized banner, the system ensures that every user is met with banners showcasing images and captions which align with their individual preferences and behaviors.

[0045] The introduction of personalized banners, where content is adapted based on the identified persona of a user, has the potential to enhance user engagement dramatically. When users encounter content that feels familiar and speaks directly to their interests, they naturally become more engaged and are more likely to interact with the respective platform (e.g. use the respective service). Moreover, personalized banners, for example, may lead to a notable increase in the Click-Through Rates (CTR) for the banners. A user experience that feels personal and relevant often prompts users to explore further, resulting in higher interactions, potentially boosting conversions and overall user satisfaction.

[0046] As users continually experience and expect personalized interactions, platforms which offer services like food delivery, e-hailing etc., which actively cater to these expectations, stand out from competitors that employ a generic content strategy. Furthermore, by continually observing user interactions with the tailored banners, deeper insights can be gained into user behaviors and preferences. This constant stream of data enables the continuous refinement of our personalization algorithms, paving the way for even more tailored user interactions in the future.

[0047] It should be noted that the image personalization system may not only be applied to banners but the principles and technology underpinning the system can be seamlessly extended to other areas of, for example, a service provider platform. This includes personalizing notifications and offering tailored content suggestions, ensuring that users experience a cohesive, customized journey from start to finish.

[0048] In essence, various embodiments allow transforming the user experience on a service provider platform from a generalized interaction to a meticulously curated personal experience.

[0049] FIG. 2 illustrates an image personalization system 200 (i.e. a system for generating personalized image data) according to an embodiment.

[0050] According to various embodiments, the process for generating personalized image data generation performed by the image personalization system 200 starts with the training of a LoRA 201 on the basis of a generic stable diffusion (SD) model 202. Given a repository of training images 213 (e.g. photographs, e.g. existing banners of a service provider) as well as illustrations and iconography, these images can be generate to create a LoRA 201 (which may be seen as an enhancement of a generic stable diffusion (SD) model 202). This allows creating images that are able to interpret the service context for which a certain image is generated, e.g. in a style commonly used for the service. For example, a driver image can be generated as such with the style that also used for other contexts of the service (e.g. using a certain color palette).

[0051] The stable diffusion network may be used to create multiple LoRAs 201 to suit different styles such as the two styles “illustrations” and “hyper-realism”, i.e. one LoRA 201 is trained for each of these two styles. This may for example be desirable because key visuals (KVs) are usually one of these two styles. The training pipeline would leverage on available illustrations to create the illustration LoRA (i.e. the LoRA for the “illustration” style) and available realistic KV images to create the hyper-realism LoRA (i.e. the LoRA for the “hyper-realism” style).

[0052] Further, each LoRA 201 is trained to recognize certain tags that are only applicable to the use case (e.g. corresponding to the service provided by the service provider for which the images are to be generated), such as drivers and passengers, merchants (categories, names, etc.), food dishes, consumer segments, etc. This is done by leveraging an available image dataset (with the desired style), which is manually annotated by merchants, consumers, and drivers etc. and fed into the training pipeline (i.e. used as training data elements for the LoRA 201). Thus, once the LoRA 201 is created, it can be used to create images that fit the respective use case, .e.g. a service provider’s context.

[0053] To use the LoRA to generate a suitable content (image), a text prompt is required to produce the image.

[0054] To create this text prompt (prompt engineering 204), an LLM is used to produce the prompt based on results from a segmentation engine 205 and a ranking engine 206. For this, to better utilize the trained (or fine-tuned) LoRA 201, a segmentation of users into user personas (generally: types of users) 207 by the segmentation engine 205 and the priority of the respective (e.g. business) use case 208, generated by the ranking engine 206, is used. This information is compiled by analyzing a database (e.g. Online Analytical Processing (OLAP) database) 209, e.g. by means of a clustering model and linear optimization. With the user personas 207, it is possible to provide the best context possible to capture the most fitting image and/or caption.

[0055] For example, the segmentation engine 205 outputs a user for which an image should be generated as “family-oriented” (i.e. assigns it to the “family-oriented” user persona which was generated by clustering of the users), and the ranking engine 206 outputs the use case as “discount program acquisition” (i.e. outputs this use case as the most important use case for the respective user). These two outputs are used by the LLM to produce the prompt that is necessary for the LoRA 201. This means that the context (user type “family-oriented” and use case “discount program acquisition”) is parsed in the prompt engineering (which includes using the LLM to generate a prompt for the LoRA 201) to generate keywords that can be used as an input to the LoRA 201 to generate a fitting image and/or caption.

[0056] So, on a high level, the image personalization system 200 can be divided into three components, the segmentation engine 205, the ranking engine 206, and an image engine 215 including the LoRA 201 (or multiple LoRAs, one of each of multiple styles) and the engineering 204 for the LoRA.

- The ranking engine automates the prioritization of business use cases such as cross-selling, up-selling, retention, and acquisition. This addresses the challenge of limited visible space in an app where elements like banners and tiles often compete. By determining the most relevant use case for display for a certain user content presentation is streamlined. The ranking is for example done by a ranking algorithm 211 which focuses on optimizing an objective function based on a system of linear equations derived from pertinent metrics 212, e.g. aiming to maximize the service provider’s objectives and key results. Given the variety of objectives the service provider might have, the ranking engine ensures that the most crucial content reaches the users, leading to optimal results.

- The segmentation engine a unique persona to every user by leveraging machine learning 214 for user segmentation. By tapping into the Online Analytical Processing (OLAP) database 209, comprehensive historical user data is collected, encompassing demographics like age and gender, transaction records such as mobility, deliveries, payments, and financial services, along with app engagement metrics like attempt rates and clickthrough rates. The mentioned categories are illustrative and not exhaustive. The collected user data is processed through one or more machine learning algorithms (in particular a clustering algorithm) to determine a distinct persona for each user. Some persona examples are “Foodie” (for those keen on foods), “Flash” (for quick service seekers), “Motorhead” (for automobile aficionados), “Travellers” (covering both inter and intra SEA journeys), among others.
- The image engine delivers tailored image results for every persona and use case, ensuring a customized experience for varied user groups and scenarios. Once validated (in a post-validation, not shown in FIG. 2), the generated images 210 are presented to the users. The image generation process employs prompt engineering, where (e.g. Python) dictionaries are used to infuse descriptions related to personas, business scenarios, and content categories into a standard prompt. This produces the individualized content fed to post-validation and (if successfully validated) displayed to the users. An image 210 may be stitched together from a background, a segmented image unique to the respective user persona, and a tagline (generally: text annotation) composed uniquely for the respective user persona.

[0057] So, integrating these three components, the OLAP database 209 is harnessed to supply the necessary data to these three components, producing tailored content (images 210) that can be presented to diverse users through the respective platform (or one or more tools thereof). This includes establishing the essential flags to categorize users and display pertinent content to them. The image personalization system 200 thus ensures the generation of evolving content and engagement landscape, adaptable to both longstanding and emerging services, products, etc., guaranteeing good user experience for all users.

[0058] In the following, an example of an image generation is given.

[0059] Step 1: by a first LLM prompt, have the LLM to describe an image that can appeal to people with a particular persona based on the business use case.

- General form of first LLM prompt: {{business_usecase}} Describe an image where {{persona}} will be interested to click onto.
- Exemplary first LLM prompt: “GrabUnlimited” is a subscription service which provides massive discounts to its subscribers. Describe an image where family-oriented people will be interested to click onto.
- Exemplary first LLM output: For an image that would attract family-oriented people to the GrabUnlimited subscription service, I envision a warm and inviting scene. It could feature a happy family of four (a mother, a father, a young daughter, and a young son) gathered around a dining table in a cozy home setting. The table is adorned with a variety of delicious-looking meals, suggesting a family dinner. Above the table, a glowing, transparent overlay of the GrabUnlimited logo is prominently displayed, symbolizing the source of these delightful meals. The family is depicted enjoying their time together, with the parents smiling at their children, who are excitedly reaching for the food. The background may include a kitchen or living room that's well-decorated, emphasizing a comfortable, familial atmosphere. The image conveys the message that GrabUnlimited can enhance family moments by providing easy access to a variety of meals, suitable for family gatherings. The overall tone of the image is warm, inviting, and relatable to anyone who values family time.

[0060] Step 2: by a second LLM prompt, have the LLM put the output of step 1 in the form of a prompt for stable diffusion

- Exemplary second LLM prompt: describe it as a prompt for stable diffusion
- Exemplary second LLM output: create an image that captures the essence of family warmth and togetherness, enhanced by the convenience of the GrabUnlimited service. In the center, depict a happy family of four - a mother and father, along with their young daughter and son, gathered around a dining table in a cozy, well-lit home setting. The table should be adorned with a variety of appetizing meals, indicating a family dinner occasion. Above this delightful scene, incorporate the transparent, glowing logo of GrabUnlimited, symbolizing the ease and variety of meal options it provides. The family is portrayed enjoying their meal and each other's company, with the parents smiling lovingly at their children who are eagerly reaching for the food. The background should include elements of a warm, inviting kitchen or living room, with tasteful decorations that emphasize a homely and comfortable environment. The

overall tone of the image should be inviting, relatable, and heartwarming, clearly conveying the message that GrabUnlimited enriches family moments with its convenient and diverse dining options.

[0061] Step 3: by a third LLM prompt, have the LLM propose a tagline

- General form of second LLM prompt: suggest an accompanying text along with the image, along the lines of {promotion}
- Exemplary second LLM prompt: suggest an accompanying text along with the image, along the lines of "Get up to $3 off every delivery. Try it for free. "
- Exemplary third LLM output: Enjoy a hearty family meal with up to $3 off. Feast now for free.

[0062] Finally, the second LLM output is used as a prompt for the LoRA 201 to generate a personalized image, and the third LLM output is used as a tagline for the personalized image, i.e. the personalized image is stitched together with the tagline and a background, then it is validated and eventually (if successfully validated) displayed to the user. For the validation, for example, a user (e.g. marketing team) ensures the final output follows predetermined design principles of the service provider, e.g. in terms of the color used, font used, and the language used. This may also include ensuring that the image generated follows stylistic guidelines of the service provider, i.e. the image is sharp and clear. This may also be performed by a corresponding software tool (e.g. a machine learning model trained to classify images into those following design principles and/or stylistic guidelines and those which do not).

[0063] In summary, according to various embodiments, a method is provided as illustrated in FIG. 3.

[0064] FIG. 3 shows a flow diagram 300 illustrating a method for generating individualized images.

[0065] In 301, a text-to-image model is trained.

[0066] In 302, by means of a large language model, a text prompt for the text-to-image model is generated by instructing the large language model to provide a text prompt for the text-to-image model for a given context including at least one of a use case and a type of user for which the individualized image is to be generated.

[0067] In 303, an individualized image for the given context is generated by feeding the generated text prompt to the text-to-image model.

[0068] According to various embodiments, in other words, an LLM is used to generate the text prompt for a text-to-image model to fit a certain context.

[0069] The method of FIG. 3 is for example carried out by a server computer as illustrated in FIG. 4.

[0070] FIG. 4 shows a server computer 400 according to an embodiment.

[0071] The server computer 400 includes a communication interface 401 (e.g. configured to receive information about the context, the specification of the text-to-image model and/or output the generated individualized image and/or configured to communicate with the LLM). The server computer 400 further includes a processing unit 402 and a memory 403. The memory 403 may be used by the processing unit 402 to store, for example, data to be processed as well as specification of the text-to-image model. The server computer is configured to perform the method of FIG. 3.

[0072] The methods described herein may be performed and the various processing or computation units and the devices and computing entities described herein may be implemented by one or more circuits. In an embodiment, a "circuit" may be understood as any kind of a logic implementing entity, which may be hardware, software, firmware, or any combination thereof. Thus, in an embodiment, a "circuit" may be a hard-wired logic circuit or a programmable logic circuit such as a programmable processor, e.g. a microprocessor. A "circuit" may also be software being implemented or executed by a processor, e.g. any kind of computer program, e.g. a computer program using a virtual machine code. Any other kind of implementation of the respective functions which are described herein may also be understood as a "circuit" in accordance with an alternative embodiment.

[0073] While the disclosure has been particularly shown and described with reference to specific embodiments, it should be understood by those skilled in the art that various changes in form and detail may be made therein without departing from the spirit and scope of the invention as defined by the appended claims. The scope of the invention is thus indicated by the appended claims and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced.

CLAIMS

1. A method for generating individualized images, comprising:
Training a text-to-image model;
Generating, by means of a large language model, a text prompt for the text-to-image model by instructing the large language model to provide a text prompt for the text-to-image model for a given context including at least one of a use case and a type of user for which the individualized image is to be generated; and
Generating an individualized image for the given context by feeding the generated text prompt to the text-to-image model.

2. The method of claim 1, wherein the type of user is a user persona.

3. The method of claim 2, wherein determining the context comprises determining the user persona from a set of user personas.

4. The method of claim 3, comprising generating the set of user personas by segmenting a set of given users.

5. The method of claim 4, comprising segmenting the set of given users by a machine learning clustering model.

6. The method of any one of claims 1 to 5, wherein the text-to-image model is a low rank adaption model of a stable diffusion model.

7. The method of claim 6, comprising generating the low rank adaption model from a predetermined stable diffusion model.

8. The method of claim 7, wherein generating the low rank adaption model comprises adapting the stable diffusion model to a predetermined image style.

9. The method of any one of claims 1 to 8, comprising generating multiple low rank adaption models, wherein the low rank adaption models are associated with different styles and the individualized image is generated by feeding the generated text prompt to the text-to-image model associated with a desired image style.

10. The method of claim 9, comprising receiving a specification of the desired image style.

11. The method of any one of claims 1 to 10, comprising receiving a specification of the context.

12. The method of any one of claims 1 to 11, comprising generating a text annotation for the individualized image and including the text annotation into the individualized image.

13. The method of claim 12, comprising generating the text annotation depending on the given context.

14. The method of claim 13, comprising generating the text annotation by instructing the large language model to provide the text annotation for the given context.

15. The method of any one of claims 1 to 14, further comprising stitching the individualized image to a predetermined background image.

16. A server computer comprising a radio interface, a memory interface and a processing unit configured to perform the method of any one of claims 1 to 15.

17. A computer program element comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of claims 1 to 15.

18. A computer-readable medium comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of claims 1 to 15.