



Bridging artificial intelligence in medicine with generative pre-trained transformer (GPT) technology

Ethan Waisberg¹, Joshua Ong², Sharif Amit Kamran³, Mouayad Masalkhi¹, Nasif Zaman³, Prithul Sarker³, Andrew G. Lee^{4,5,6,7,8,9,10,11}, Alireza Tavakkoli³

¹University College Dublin School of Medicine, Belfield, Dublin, Ireland; ²Michigan Medicine, University of Michigan, Ann Arbor, MI, USA; ³Human-Machine Perception Laboratory, Department of Computer Science and Engineering, University of Nevada, Reno, NV, USA; ⁴Center for Space Medicine, Baylor College of Medicine, Houston, TX, USA; ⁵Department of Ophthalmology, Blanton Eye Institute, Houston Methodist Hospital, Houston, TX, USA; ⁶The Houston Methodist Research Institute, Houston Methodist Hospital, Houston, TX, USA; ⁷Departments of Ophthalmology, Neurology, and Neurosurgery, Weill Cornell Medicine, New York, NY, USA; ⁸Department of Ophthalmology, University of Texas Medical Branch, Galveston, TX, USA; ⁹University of Texas MD Anderson Cancer Center, Houston, TX, USA; ¹⁰Texas A&M College of Medicine, Bryan, TX, USA; ¹¹Department of Ophthalmology, The University of Iowa Hospitals and Clinics, Iowa, IA, USA

Correspondence to: Ethan Waisberg, MB, BCh, BAO. University College Dublin School of Medicine, Belfield, Dublin 4, D04 C1P1, Ireland. Email: ethan.waisberg@ucdconnect.ie.

Abstract: Since its public release in November 2022, the usage of ChatGPT (Open AI, USA) has been unprecedented. This large language model (LLM) can produce human-like text from deep-learning techniques. LLMs are rapidly approaching human-level performance. ChatGPT can potentially help democratize the ability to code, by allowing clinicians to be able to develop basic artificial intelligence (AI) techniques. By leveraging AI models, these clinicians can expand the scope of their research abilities, and this can potentially lead to an AI in medicine revolution, where clinicians are able to generate clinically-focused AI techniques with the goal of improving patient outcomes across all domains. In this paper, we examine the performance of ChatGPT at developing an AI program for medicine and its associated limitations and challenges. Similar to the majority of AI models, the ethical concerns surrounding its application in medicine remains, which includes biases, patient autonomy, and confidentiality, transparency, and accuracy of data. ChatGPT must also be used in accordance with local healthcare regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States. All things considered, ChatGPT and future generative AI technologies will democratize the ability to code and develop AI, likely leading to breakthroughs in the medical AI sector.

Keywords: Artificial intelligence (AI); medicine; ChatGPT; large language model (LLM)

Received: 27 April 2023; Accepted: 20 July 2023; Published online: 01 August 2023.

doi: 10.21037/jmai-23-36

View this article at: <https://dx.doi.org/10.21037/jmai-23-36>

Since its public release in November 2022, the usage of ChatGPT (Open AI, USA) has been unprecedented. This large language model (LLM) can produce human-like text from deep-learning techniques. Generative pre-trained transformer (GPT) architecture, processes and generates text using a transformer neural network. As the size of the training dataset increases, the LLM can better infer relationships of words and produce more human-like text.

This generative model has already been capable

to do a variety of useful medical tasks, from writing discharge summaries (1) to generating images from patient descriptions of neuro-ophthalmic conditions (2) to helping with triaging of ophthalmic conditions (3). LLMs are rapidly approaching human-level performance, with ChatGPT successfully completing the Royal College of General Practitioners Applied Knowledge Test with an average score of 60.17% (4). In another recent study, ChatGPT was shown to be able to respond to patient

```
python
import tensorflow as tf
from tensorflow.keras import layers, models

# Define the model architecture
model = models.Sequential([
    layers.Conv2D(32, (3, 3), activation='relu', input_shape=(224, 224, 3)),
    layers.MaxPooling2D((2, 2)),
    layers.Conv2D(64, (3, 3), activation='relu'),
    layers.MaxPooling2D((2, 2)),
    layers.Conv2D(128, (3, 3), activation='relu'),
    layers.MaxPooling2D((2, 2)),
    layers.Flatten(),
    layers.Dense(128, activation='relu'),
    layers.Dropout(0.5),
    layers.Dense(4, activation='softmax')
])

# Compile the model
model.compile(optimizer='adam',
              loss='categorical_crossentropy',
              metrics=['accuracy'])

# Train the model
history = model.fit(X_train, y_train, epochs=10, batch_size=32,
                   validation_data=(X_test, y_test))

# Evaluate the model
```

Figure 1 From the prompt “create a CNN with Keras for OCT classification for normal, glaucoma, drusen and CNV classes”. CNN, convolutional neural network; OCT, optical coherence tomography; CNV, choroidal neovascularization.

questions from a social media forum with higher levels of empathy and quality than the responses provided by physicians (5). ChatGPT can also be able to generate basic code, which can be useful for clinician-scientists to begin working with artificial intelligence (AI) for the first time.

We are a multidisciplinary healthcare and engineering team working on developing solutions to monitor and maintain astronaut vision during long-duration spaceflight (LDSF) (6-9). Although highly beneficial, multidisciplinary collaborations can also give rise to a multitude of challenges including organization and communication which results in delays and decreases in productivity. The usage of ChatGPT can help democratize the ability to code, by allowing clinicians to be able to develop basic AI techniques. By leveraging AI models, these clinicians can expand the scope of their research abilities, and this can potentially lead to AI in medicine breakthroughs, where clinicians are able to generate clinically-focused AI techniques with the goal of improving patient outcomes across all domains.

To examine the performance of ChatGPT at developing an AI program for medicine, we asked ChatGPT to develop a convolutional neural network (CNN) to classify optical coherence tomography (OCT) images in ophthalmology

(*Figure 1*). All responses in this paper were generated with GPT-3.5 in early March, 2023 prior to the release of GPT-4 (10). According to ChatGPT, the CNN it produced is capable of predicting probabilities of a normal, glaucoma, drusen and choroidal neovascularization (CNV) OCT. We then had an expert in machine learning grade ChatGPT’s response.

The CNN produced by ChatGPT to classify OCTs was already quite advanced with 3 convolutional layers, 2 dense layers, and max pooling, but could be improved upon. To improve upon this, we then asked ChatGPT to include normalization and residual connections (*Figure 2*). The code was then successfully updated with batch normalization after every convolutional layer which normalized the feature maps. Skip connections were also added by ChatGPT to help the network learn residual functions.

We then asked ChatGPT to further make improvements to this code by introducing a learning rate reduction strategy, so that the network does not get stuck in local minima (*Figure 3*). ChatGPT was again able to successfully alter the code, and used a strategy which involves reducing the learning rate if the validation loss has not improved over a certain number of epochs.

```

import tensorflow as tf
from tensorflow.keras import layers, models

# Define the model architecture
inputs = layers.Input(shape=(224, 224, 3))

x = layers.Conv2D(32, (3, 3), padding='same')(inputs)
x = layers.BatchNormalization()(x)
x = layers.Activation('relu')(x)
x = layers.MaxPooling2D((2, 2))(x)

residual = x

x = layers.Conv2D(64, (3, 3), padding='same')(x)
x = layers.BatchNormalization()(x)
x = layers.Activation('relu')(x)
x = layers.Conv2D(64, (3, 3), padding='same')(x)
x = layers.BatchNormalization()(x)
x = layers.Add()([x, residual])
x = layers.Activation('relu')(x)
x = layers.MaxPooling2D((2, 2))(x)

residual = x

x = layers.Conv2D(128, (3, 3), padding='same')(x)
x = layers.BatchNormalization()(x)
x = layers.Activation('relu')(x)
x = layers.Conv2D(128, (3, 3), padding='same')(x)
x = layers.BatchNormalization()(x)
x = layers.Add()([x, residual])
x = layers.Activation('relu')(x)
x = layers.MaxPooling2D((2, 2))(x)

x = layers.Flatten()(x)
x = layers.Dense(128, activation='relu')(x)
x = layers.Dropout(0.5)(x)
outputs = layers.Dense(4, activation='softmax')(x)

model = models.Model(inputs=inputs, outputs=outputs)

# Compile the model
model.compile(optimizer='adam',
              loss='categorical_crossentropy',
              metrics=['accuracy'])

# Train the model
history = model.fit(X_train, y_train, epochs=10, batch_size=32,
                   validation_data=(X_test, y_test))

# Evaluate the model
test_loss, test_acc = model.evaluate(X_test, y_test)
print('Test accuracy:', test_acc)

```

Figure 2 Generated from the prompt “Can you improve this CNN with normalization and residual connections”. CNN, convolutional neural network.

The final task we used to improve the generated CNN was by implementing a data augmentation strategy (Figure 4). Data augmentation can improve the performance of a CNN by increasing training data diversity and reduce overfitting.

ChatGPT displays a highly promising ability to

develop and refine code to create useful CNN models. ChatGPT can be a promising tool to help bridge the gap and allow physicians to develop novel applications of AI. It is important to consider limitations of large language learning models, which can potentially make errors and also

```

import tensorflow as tf
from tensorflow.keras import layers, models, callbacks

# Define the model architecture
inputs = layers.Input(shape=(224, 224, 3))

x = layers.Conv2D(32, (3, 3), padding='same')(inputs)
x = layers.BatchNormalization()(x)
x = layers.Activation('relu')(x)
x = layers.MaxPooling2D((2, 2))(x)

residual = x

x = layers.Conv2D(64, (3, 3), padding='same')(x)
x = layers.BatchNormalization()(x)
x = layers.Activation('relu')(x)
x = layers.Conv2D(64, (3, 3), padding='same')(x)
x = layers.BatchNormalization()(x)
x = layers.Add([x, residual])
x = layers.Activation('relu')(x)
x = layers.MaxPooling2D((2, 2))(x)

residual = x

x = layers.Conv2D(128, (3, 3), padding='same')(x)
x = layers.BatchNormalization()(x)
x = layers.Activation('relu')(x)
x = layers.Conv2D(128, (3, 3), padding='same')(x)
x = layers.BatchNormalization()(x)
x = layers.Add([x, residual])
x = layers.Activation('relu')(x)
x = layers.MaxPooling2D((2, 2))(x)

x = layers.Flatten()(x)
x = layers.Dense(128, activation='relu')(x)
x = layers.Dropout(0.5)(x)
outputs = layers.Dense(4, activation='softmax')(x)

model = models.Model(inputs=inputs, outputs=outputs)

# Define the learning rate reduction strategy
reduce_lr = callbacks.ReduceLROnPlateau(monitor='val_loss', factor=0.1,
                                         patience=5, verbose=1, min_delta=1e-4)

# Compile the model
model.compile(optimizer='adam',
              loss='categorical_crossentropy',
              metrics=['accuracy'])

# Train the model
history = model.fit(X_train, y_train, epochs=10, batch_size=32,
                   validation_data=(X_test, y_test), callbacks=[reduce_lr])

# Evaluate the model

```

Figure 3 Generated from the prompt “Can you further improve this by introducing reduce learning rate on plateau so that the loss does not get stuck in local minima”.

unintentionally plagiarize work (11). All things considered, ChatGPT and future generative AI technologies will democratize the ability to code and develop AI, likely leading to breakthroughs in the medical AI sector. GitHub has recently introduced “Copilot Chat”, a built-in ChatGPT-like experience to help coders by providing in-

depth explanations and analysis.

On the other hand, ChatGPT also presents with its unique set of challenges and limitations that are important to recognize. Similar to the majority of AI models, the ethical concerns surrounding its application in medicine remains, which includes biases, patient autonomy, and confidentiality,

```
python
import tensorflow as tf
from tensorflow.keras import layers, models, callbacks, optimizers
from tensorflow.keras.preprocessing.image import ImageDataGenerator

# Define the data generators with data augmentation
train_datagen = ImageDataGenerator(rescale=1./255,
                                   rotation_range=20,
                                   width_shift_range=0.2,
                                   height_shift_range=0.2,
                                   shear_range=0.2,
                                   zoom_range=0.2,
                                   horizontal_flip=True)

test_datagen = ImageDataGenerator(rescale=1./255)

train_generator = train_datagen.flow(X_train, y_train, batch_size=32)
test_generator = test_datagen.flow(X_test, y_test, batch_size=32)
```

Figure 4 Generated with the prompt “Can you further improve this model by introducing a data augmentation strategy for training the model”.

transparency, and accuracy of data. ChatGPT must also be used in accordance with local healthcare regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States. To ensure this is occurring, patient protected health information must be stored and transmitting securely, while following strict authentication protocols. Compliance with these regulations must also be regularly assessed. To guarantee that ChatGPT is utilized in a responsible manner, these issues must be addressed. The volume of data that ChatGPT is prepared using has a major effect on the precision and reliability of its suggestions. This is especially important in medicine; ChatGPT may not fully ‘understand’ the complexities and nuances of medical terminology, potentially leading to inaccurate feedback. In conclusion, even though ChatGPT has the potential to alter healthcare, its application in medicine has to be closely watched and assessed to make sure it is utilized in an ethical and responsible way.

Acknowledgments

Funding: This study was supported by NASA Grant (No. 80NSSC20K183): A Non-intrusive Ocular Monitoring Framework to Model Ocular Structure and Functional Changes due to Long-term Spaceflight.

Footnote

Peer Review File: Available at <https://jmai.amegroups.com/article/view/10.21037/jmai-23-36/prf>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://jmai.amegroups.com/article/view/10.21037/jmai-23-36/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health* 2023;5:e107-8.
2. Waisberg E, Ong J, Masalkhi M, et al. Text-to-image artificial intelligence to aid clinicians in perceiving unique neuro-ophthalmic visual phenomena. *Ir J Med Sci* 2023. [Epub ahead of print]. doi: 10.1007/s11845-023-03315-8.
3. Waisberg E, Ong J, Zaman N, et al. GPT-4 for triaging ophthalmic symptoms. *Eye (Lond)* 2023. [Epub ahead of

- print]. doi: 10.1038/s41433-023-02595-9.
4. Thirunavukarasu AJ, Hassan R, Mahmood S, et al. Trialling a Large Language Model (ChatGPT) in General Practice With the Applied Knowledge Test: Observational Study Demonstrating Opportunities and Limitations in Primary Care. *JMIR Med Educ* 2023;9:e46599.
 5. Ayers JW, Poliak A, Dredze M, et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Intern Med* 2023;183:589-96.
 6. Waisberg E, Ong J, Zaman N, et al. Head-Mounted Dynamic Visual Acuity for G-Transition Effects During Interplanetary Spaceflight: Technology Development and Results from an Early Validation Study. *Aerosp Med Hum Perform* 2022;93:800-5.
 7. Waisberg E, Ong J, Paladugu P, et al. Optimizing screening for preventable blindness with head-mounted visual assessment technology. *J Vis Impairment Blindness* 2022;116:579-81.
 8. Ong J, Zaman N, Waisberg E, et al. Head-mounted digital metamorphopsia suppression as a countermeasure for macular-related visual distortions for prolonged spaceflight missions and terrestrial health. *Wearable Technol* 2022;3:e26.
 9. Ong J, Tavakkoli A, Zaman N, et al. Terrestrial health applications of visual assessment technology and machine learning in spaceflight associated neuro-ocular syndrome. *NPJ Microgravity* 2022;8:37.
 10. Waisberg E, Ong J, Masalkhi M, et al. GPT-4: a new era of artificial intelligence in medicine. *Ir J Med Sci* 2023. [Epub ahead of print]. doi: 10.1007/s11845-023-03377-8.
 11. Alser M, Waisberg E. Concerns with the usage of ChatGPT in Academia and Medicine: A viewpoint. *Am J Med Open* 2023;9:100036.

doi: 10.21037/jmai-23-36

Cite this article as: Waisberg E, Ong J, Kamran SA, Masalkhi M, Zaman N, Sarker P, Lee AG, Tavakkoli A. Bridging artificial intelligence in medicine with generative pre-trained transformer (GPT) technology. *J Med Artif Intell* 2023;6:13.