

Artificial Intelligence and Large Language Models in the Fight Against Superficial Fungal Infections: Friend or Foe?

Aditya K Gupta^{1,2}, Vasiliki Economopoulos²

¹Division of Dermatology, Department of Medicine, University of Toronto, Toronto, Ontario, Canada; ²Mediprobe Research Inc., London, Ontario, Canada

Correspondence: Aditya K Gupta, Mediprobe Research Inc., 645 Windermere Road, London, Ontario, N5X 2P1, Canada, Tel +1-519-851-9715, Fax +1-519-657-4233, Email agupta@mediproberesearch.com

Abstract: Superficial fungal infections can have significant physical and psychological consequences for affected patients. These painful infections have become more prevalent and the rise of antifungal resistant strains is of great concern. New tools in the fight against these infections are needed, especially in areas where appropriate dermatological care is lacking. Artificial intelligence (AI) offers a potential solution for these care gaps. AI's capabilities have been increasing in sophistication at an astonishing pace, with large language models (LLMs), such as ChatGPT (OpenAI), Claude (Anthropic) and Gemini (Google) being capable of generating detailed responses to complex problems as well as demonstrating reasoning type behaviour. AI is currently in use and being developed for use within the clinic as well as the laboratory, with the potential to significantly improve access to dermatological care and patient outcomes. However, understanding how these AI models work at a basic level is necessary for safe, effective and efficient use and application to the management of superficial fungal infections. In this review, we provide a high-level description of how these models work, discuss the potentials and pitfalls of AI and LLMs, as well as their applications and the current and future outlook for the field.

Keywords: artificial intelligence in dermatology, superficial fungal infections, dermatology specific LLMs, ethics of artificial intelligence in medicine

Introduction

Superficial fungal infections, including onychomycosis, are infections that can not only cause significant discomfort in patients, but also create cosmetic issues that further impact patients' lives.¹ These infections preferentially affect the elderly, but also impact regions of lower socio-economic development more frequently.² These infections can be difficult to treat, having high failure rates and requiring lengthy courses of antifungal agents.³ The rate of antifungal resistant infections is on the rise and has significant implications for dermatology, particularly in less developed regions where access to adequate dermatological care may be limited. The emergence of terbinafine-resistant strains, such as *Trichophyton indotineae* underscores the need to monitor patient care.⁴

With the clinical challenges that we are facing in managing superficial fungal infections, there is a need for new tools that can assist physicians in the diagnosis and treatment of these infections. Tools that can hasten the identification of pathogens and assist in determining drug susceptibility will improve treatment outcomes and help to reduce the risk of new anti-fungal resistant strains from emerging. Artificial intelligence (AI) platforms, and in particular large language models (LLMs), have the potential to create these types of tools, leading to better patient outcomes, reduced physician workload and improved laboratory throughput. Several LLMs have been developed and are available commercially for general use, including OpenAI's ChatGPT (<https://chatgpt.com/>), Google's Gemini (<https://gemini.google.com/>), and Anthropic's Claude (<https://claude.ai/>), as well as many others. Our goal in this review is to provide an overview of the current and future applications of LLMs in combating superficial fungal infections,



while discussing their inherent limitations and potential dangers, providing the physician with the necessary knowledge to use these models safely and effectively.

We will cover the basics of AI and how LLMs actually work/think, the advantages, disadvantages and safety concerns these models pose, as well as discuss the current and future applications of LLMs in the diagnosis and treatment of superficial fungal infections.

What Is Artificial Intelligence?

AI-based technologies have already been deployed within medicine with great benefit, for example Apple's atrial fibrillation detection using the ECG application on the Apple Watch, Medtronic's Guardian Sugar.IQ system, and Paige.ai's AI-based pathology tools.⁵⁻⁷ However, before we can learn how to best use AI deep learning technologies such as LLMs, we must first gain an understanding of what this technology is and how it actually works.

We do not need extreme detail, but a general understanding will suffice for this review. We can think of AI systems, particularly machine learning and deep learning systems, as statistical models of the real world (Figure 1).⁸ These models use iterative statistical methods to essentially "learn" or "perceive" the relationship between a set of independent variables and a dependent variable (training).^{8,9} These variables can be a measurement, a group of measurements (or properties), or even pieces of text and images, which are relevant to LLMs. Once these relationships are determined, the model can then be used to make inferences about the value or property of the dependent variable when only given the independent variables as an input.

As a simple example, a model can be trained on data that shows that 80% of people that like oranges also like grapefruit and that 30% of people that like apples also like grapefruit. Once trained, the model can be shown data on two individuals, one that likes oranges and one that likes oranges *and* apples, and asked to predict whether these individuals will like grapefruit.

There are many different methods for developing machine/deep learning models; however, these are outside the scope of this review. We would like to refer the inclined reader to additional resources to learn more (Introduction to Deep Learning by Eugene Charniak; Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville). Understanding the basic principles discussed above will help us understand how LLMs work and how to best work with these models.

What Is a Large Language Model?

LLMs have a seemingly astonishing capacity to interpret unstructured information that the user provides and craft a response that is not only informative, but also accurate and useful. However, these models are once again only statistical models of the real world. Language models – and LLMs by extension – require natural language to be broken into small pieces of information through a process called tokenization, where characters, words and/or portions of words are assigned unique numerical values or "tokens" (Figure 2). These models are then provided with a sequence of tokens which represents the sentences entered by the user. Similar to AI models in general, these language models only truly predict the next most likely "word" or token in the response based on the sequence of information that was given.⁹ LLMs are trained on vast amounts of text data, which consists vetted information extracted from internet sources, in order to achieve these feats.^{10,11} Training these models can be costly with the need for large amounts of computing power as the number of parameters that need to be optimized are extremely high – for example OpenAI's GPT-3.5 has approximately 175 billion parameters.^{10,12}

Training also occurs in multiple steps, with the first step training the LLM to predict the next token in a sequence forming its knowledge base. The second step is to fine tune the model so that it can respond to instruction, and the third is to further fine tune the model to produce desired behaviour through reinforcement learning. It is important to note that the date that the LLM was trained will impact its knowledge base. An LLM will only have "memory" for information that was present by the training dataset's cut-off date, if asked about information that would not have been in the training dataset (eg a medication that was approved a year after the cut-off date, or a newly emergent drug-resistant strain), it will not be able to provide accurate information.

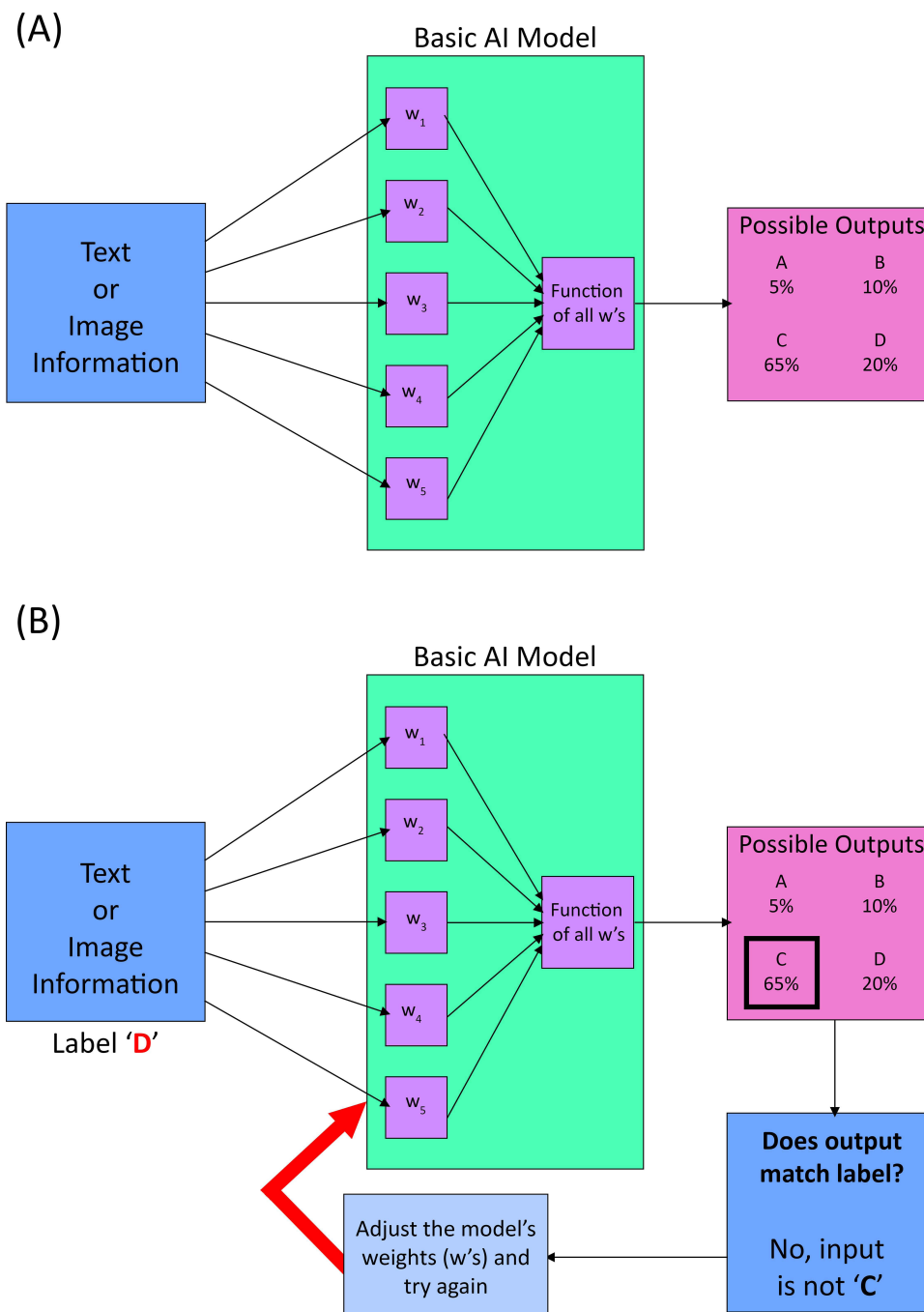


Figure 1 Simplified model of artificial intelligence (AI) system. The AI model, in this case a very simple neural network (A), receives parametrized information of either a piece of text or image and performs an initial calculation with the weights shown in the purple boxes. The model then conducts a second calculation with the results of the first, obtaining probabilities that the information is in a certain category. During training (B) the model is fed the same type of information and calculates the most likely category. However, the model then compares its answer to the ground truth category (or label) of the information. If the model's answer and the information's ground truth label do not match, then the weights of the model are adjusted slightly and the process is run again.

AI and LLMs: Potential and Pitfalls

With the advent of more powerful chips designed to handle the demands of AI technologies, but with higher efficiency and lowered costs, as well as the increasing amounts of online data available for training and advanced algorithms, the potential for developing sophisticated tools to assist the dermatologist is staggering. Koo et al have developed a deep learning model to identify the presence of fungal hyphae in potassium hydroxide (KOH) stains of skin and nail scrapings, and Zieliski et al have developed a model to specifically identify fungal species from Gram stained samples.^{13,14}

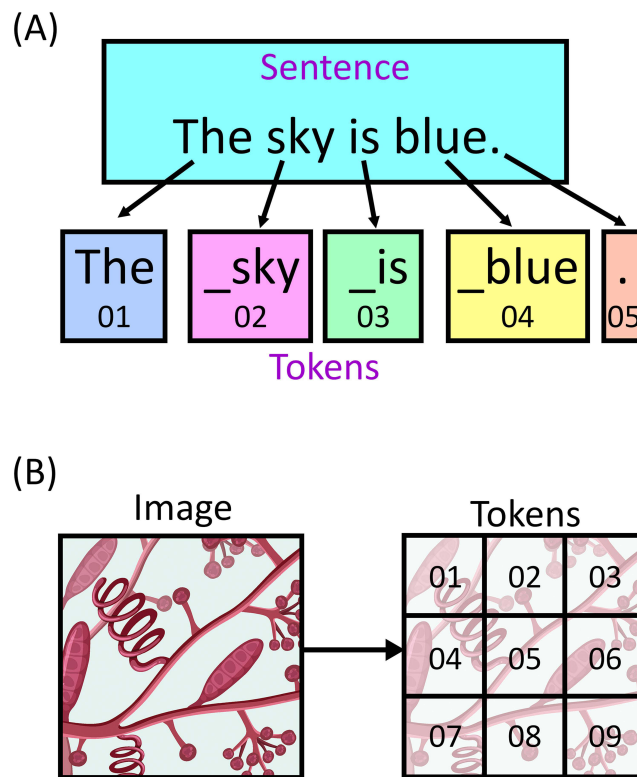


Figure 2 Tokenization of data for large language models. **(A)** The tokenization of text data. The sentence which consists of 4 words, 3 spaces and 1 punctuation mark is split into distinct “tokens”. These tokens are each represented by a unique code and can consist of a combination of words, parts of words, spaces with words and punctuation marks. **(B)** Image data can also be tokenized where the information within the image is split into smaller individual pieces that can then be passed to the model.

However, there are some significant concerns surrounding the use of AI and LLMs in medical practice. These include, but are not limited to cost, accuracy, and regulatory and ethical issues.¹⁵ Understanding the trade-offs is essential for health care professionals to effectively and safely use these tools. Below we discuss the potentials as well as the pitfalls that users of these technologies need to be aware of, in addition to listing in [Table 1](#).

How AI Models Can Help the Physician

The applications mentioned above are only some of the possible applications of AI system. The very recently developed emGene system, which utilizes a specially trained LLM installed locally on a next-generation sequencer (NGS) system, could potentially be used for fungal species identification and anti-fungal resistance gene detection, and provide reports to physicians who may not have the knowledge or experience to interpret raw gene sequencing data.¹⁶ These types of systems have the potential to not only decrease physician workload, but also improve consistency and make this type of testing, particularly genetic tests, more accessible.

These models also have the potential to greatly improve clinical workflows, and not only for the analysis of laboratory samples as mentioned above. Specially trained LLMs, such as Med-Alpaca¹⁷ and PMC-LLaMA,¹⁸ with their vast medical knowledge base, have the potential to act as consultation tools.¹⁹ These models could be used by the physician to aid in clinical decisions. For example, after providing information about the patient’s skin or nail symptoms as well as the patient’s characteristics and health history, the model can provide feedback on the likelihood that the patient’s complaint is due to a fungal infection vs a bacterial infection vs a non-infectious disorder, with [Figure 3](#) highlighting these potential applications and workflows. These models can help guide testing and treatment decisions to not only achieve accurate diagnoses faster, but also help reduce the need for unnecessary testing, while promoting anti-microbial stewardship by reducing the unnecessary prescribing of antifungal drugs. This could be of particular importance in areas where specialized dermatological care is not readily available or accessible. These tools could allow general practitioners

Table 1 Advantages/Potentials and Disadvantages/Pitfalls of AI and LLMs in Managing Superficial Fungal Infections

Potentials	Pitfalls
<p><i>Pathology Settings</i></p> <ul style="list-style-type: none"> Increased throughput of samples AI assisted analysis of samples Report generation in genetic testing for species identification and detection of resistance mutations <p><i>Patient Screening</i></p> <ul style="list-style-type: none"> Differentiate infectious vs non-infectious conditions Reduce need for unnecessary testing Reduce the need for unnecessary antifungal prescriptions Promote anti-microbial stewardship <p><i>Access to Dermatological Care</i></p> <ul style="list-style-type: none"> Consultation tool to assist non-dermatologists in providing care Formation of treatment plans personalized to each patient 	<p><i>Model Accuracy</i></p> <ul style="list-style-type: none"> Can be prone to hallucinations/confabulations Quality of training data can impact accuracy Falsehood mimicry due to accepting false information from user as fact <p><i>Algorithmic errors</i></p> <ul style="list-style-type: none"> User experience in prompting Potential for patient harm Stale-dating of training data Data bias <p><i>Regulatory and Ethical Concerns</i></p> <ul style="list-style-type: none"> Breach of patient privacy Unauthorized use of patient information Regulatory status and implications Process transparency <p><i>Legal Concerns</i></p> <ul style="list-style-type: none"> Who is responsible party in event of medical error <p><i>Access Concerns</i></p> <ul style="list-style-type: none"> Cost of specialized hardware Cost of running the model Availability of sufficient infrastructure Adequate training for users

with more limited knowledge and experience with these infections to provide appropriate care and therefore improving patient outcomes.

The Downside of AI in the Clinic

The potential that these models hold is counter-balanced by many pitfalls. Currently, there are concerns around (1) model accuracy, (2) regulation and ethics, (3) legal implications and (4) access.

Model Accuracy

LLMs are trained on vast amounts of data from mainly online sources.¹¹ Even though considerable effort has been taken to vet this data for accuracy, the sheer volume of data precludes complete accuracy. This is partly due to unavoidable errors when screening such an enormous dataset, but also due to the nature of the information that would be available. Many online sources may contain information that is contradictory or misleading but may appear to be credible and even biased information. When this information is not recognized and allowed to remain in the training data, then these inaccuracies will essentially be “learned” by the model.

Another possible reason for inaccuracies may be the cut-off date of the training data. As these models are quite costly and time consuming to train and vetting the training data is also time consuming,²⁰ the dataset that is fed to the only contains information up to a specific date. As development of a new model may take many months, the information contained within the training data can become outdated. This is of particular concern in when applied to superficial fungal infections as the information surrounding the emergence of new and/or antifungal resistant strains can change extremely quickly. This type of data inaccuracy will lead to inevitable errors.

Another concern is LLM “hallucinations” or confabulations.²¹ These models truly only predict the next tokens (or series of words) in a given sequence. If the model has had very little experience through training on similar

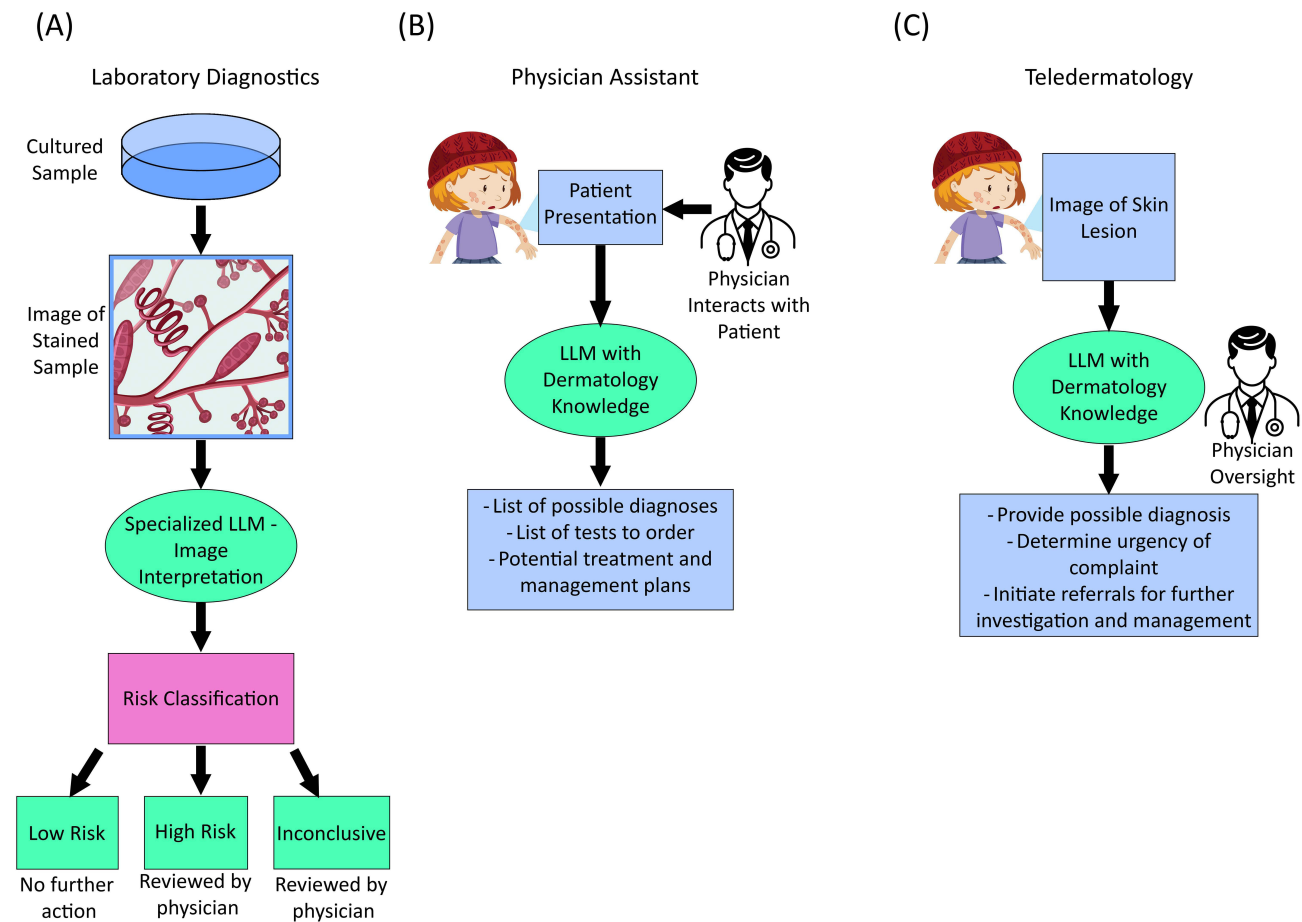


Figure 3 Applications of artificial intelligence to the management of superficial fungal infections. **(A)** When applied to laboratory diagnostic testing, a specialized AI model can interpret the image of a stained sample and then determine if there is any evidence of fungus within the image. The model could then provide a classification of the sample as either low risk, high risk, or inconclusive. Samples that are high risk or inconclusive would be reviewed by an expert to confirm diagnosis. **(B)** Large language models could also act as an assistant to the physician. The patient would be seen by the physician, who would then consult with an LLM fine-tuned with dermatology data, providing details of the patient's presentation. The model could then provide differential diagnoses, recommend appropriate diagnostic testing and assist in developing a treatment plan. **(C)** In teledermatology, a similar process could be used, where in place of the physician, the patient would communicate with the LLM and provide details and photographs of their complaint. The LLM could then either provide possible diagnoses, determine if urgent care is needed or even initiate referrals for further investigations and management. The LLM in this case would have physician oversight to ensure accuracy.

sequences (eg not trained specifically on medical knowledge),²² then it will not be able to accurately predict the next set of tokens, leading to erroneous responses. Physicians using these models must not take the output at face value.^{21,23} Additionally, these models cannot understand the difference between true and false information. If false information is inadvertently fed into the model during use, either by unintentional data entry error or vague prompting, the model can actually take the false information and mimic it back with additional information that will be incorrect.

These inaccuracies, if not recognized in a timely manner, can lead to significant patient harm. For example, a patient may be harmed by an LLM recommending an antifungal medication that is contraindicated due to the patient's medical history, or an LLM misinterprets pathology results; this may result in a patient with a clear infection not being promptly treated. This risk can be reduced by ensuring that the LLM's training data has detailed and accurate medical knowledge at all stages of training.

Regulatory and Ethical Concerns

Protecting patient's privacy and maintaining the security of their personal health information is a must. Any implementation of AI technology must uphold these principles. One of the main issues surrounding patient privacy is the requirement

for high performance hardware to run many of these models. In the case of commercial models such as ChatGPT, the model itself is hosted by the developer (in this case OpenAI) and interactions with the LLM are through the developer's provided interface and through their infrastructure. This flow of data to the developer can have significant consequences. Firstly, any patient information that is transferred online has the potential to be intercepted by malicious entities.²³ Second, many of these developers will store the information provided within the chats that the physician has with the LLM, leading to patient information being given to a third party without the explicit consent of the patient.^{11,23} Thirdly, some developers may use these stored chats as future training data unless the user specifically tells the LLM not to hold the data and use it for training. For example, OpenAI states on their website that conversations can be retained and used for training unless the user disables this option.²⁴ These issues could be potentially be rectified by hosting these models locally; however, this would require additional computing resources as well as frequent on-site updates to the model to ensure accuracy, particularly in microbiology and infectious disease applications.

We also need to consider if the use of AI algorithms and LLMs, including those such as ChatGPT, Gemini and Claude, in healthcare could be considered the use of a medical device. Currently, there are laws within both the European Union and the US that do consider AI technologies to be medical devices when developed for such purposes.²² However, commercial general use LLMs have not been built for those purposes and would not necessarily conform to these regulations. In this case, how would we classify the act of a physician consulting an LLM regarding a diagnosis or whether to collect a specimen for testing? Physicians must undergo extensive training and licencing before they can practice independently. If LLMs are to be used in a similar, yet albeit at a reduced capacity, then regulation of these systems for transparency is a must. Similar considerations also apply for LLMs that are used in the analysis of laboratory and pathology samples. Currently, it is not known how often these models, particularly those such as ChatGPT, Gemini and Claude are used in the clinical setting. However, a study of physician performance augmented by LLMs has hinted that some physicians may already be using these tools to some extent within their practice.²⁵

Legal Concerns

Inevitably, there will be medical errors made with the use of LLMs. The question then becomes: who are the responsible parties? Traditionally it would just be the practitioners and their organizations that were involved in the error that would be liable. However, with an LLM, depending on the particular use case, would the practitioner (physician, laboratory technician, nurse, hospital, etc.) be fully responsible, or are the developers of the LLM fully responsible, or is it reasonable for fault to be shared by multiple parties.²³ The legal ramifications of LLMs in practice are not clear, but will need to be seriously considered by any practitioner or organization looking to implement these models into their practice.²³

Ease of Access

As mentioned earlier, LLMs can be extremely costly. Specialized hardware, extremely large amounts of vetted training data, long training times and high energy needs all lead to high costs. In less socio-economically developed areas the cost of these technologies may be prohibitive, even though these areas would stand to benefit the most. Additionally, remote areas which may not have the electrical and internet infrastructure may also have limited access to these technologies.¹⁵ Training users appropriately in these remote areas will be more difficult and also contribute to reduced access from the perspective of the patient.

Applications to Superficial Fungal Diseases

One of the major applications of AI to superficial fungal diseases is diagnostic assistance. With the ability of LLMs to analyse different types of images, these models have potential to assist in reducing physician work load by classifying high and low risk samples, and to also potentially act as an "expert" in areas where knowledgeable practitioners may not be readily available.

As mentioned earlier, several groups have been able to develop technologies for the identification of superficial fungal infections in pathology samples. Koo et al have developed a convolutional neural network model to detect fungal hyphae in KOH stained samples.¹³ The advantage of analysing these types of images is that KOH staining is relatively quick and inexpensive compared to other testing methods. They demonstrated a high degree of accuracy in detection using data

from actual patient samples, however, the amount of data for training and testing was limited, which would inevitably lead to biases and inaccuracies when applied to images from outside of their study.

Zieliski et al have also conducted a similar study, but with the aim of identifying the species present within Gram-stained samples.¹⁴ The training set used within this study prevented the use of standard neural network architectures, but the authors were able to use different methods (bag-of-words followed by support vector machine classification) to achieve classification with a high degree of accuracy, with up to 100% accuracy for some species.¹⁴ It is important to note with this work that the available training data was much smaller than the study by Koo et al. However, even with these limitations, both of these studies demonstrate the feasibility of applying these models to screening methods.

Jansen et al have presented a deep learning system that utilizes Periodic Acid Schiff (PAS) stained whole slide images produced by commercially available histopathology slide scanners to detect fungal elements in nail samples with accuracy similar to that of trained pathologists.²⁶ Decroos et al conducted a similar study of PAS stained whole slide images, again demonstrating similar overall accuracy to dermatopathologists in side-by-side comparisons, where their model demonstrated greater accuracy than half of assessed dermatopathologists.²⁷ Even though their algorithms shows very promising results, the image artifacts that can occur with whole slide images do have the potential to lead to errors.^{26,27} Nonetheless, their work is an important demonstration of how these AI technologies can potentially fit into current clinical workflows.

Another application is identification of fungal infections in photographs of nails. This could be particularly useful in settings where access to specialized testing may be limiting, helping to spare those limited resources. One example is the work presented by Han et al in which they demonstrate a serial series of convolutional neural networks that performs with the same accuracy as dermatologists in identifying onychomycosis from photographs of nails.²⁸ One downside of this work is that the training and validation sets are from non-standardized images, adding additional variation to the dataset.²⁸ Another is that the images in the training set were derived from patients in South Korea, which may introduce biases into the model in terms of using primarily South Korean patients to train from. Kim et al then demonstrated efficacy of this algorithm to identify onychomycosis in nail photographs in a study of 90 patients independent of the training data, where they showed equivalent performance to five board-certified dermatologists.²⁹

Abdulhadi et al conducted a head-to-head comparison of five different transfer learning models (a type of LLM) in classifying nail images into either healthy, hyperpigmented, nail clubbing or onychomycosis, with the DenseNet201³⁰ model demonstrating the highest efficacy.³¹ Elbes et al have conducted a similar study, but have examined the ability of various LLMs (VGG-16/19, Inception V3, ResNet50, Inception-ResNet-v2 and MobileNet) to specifically differentiate between 11 different skin pathologies, again demonstrating efficacy.³²

A dermatology-specific LLM named SkinGPT-4 has also been developed by Zhou et al with the goal of providing diagnoses for a wide variety of skin conditions, including superficial fungal infections.³³ In this work, the authors use the MiniGPT-4 LLM as a starting point and further fine tune the model with dermatology-specific data, leading to a potentially effective diagnostic tool in areas with reduced dermatological care.³³ However, there is the potential for errors which could be harmful to patients in situations where physician oversight is not available.

The work of Han et al, Kim et al, Abdulhadi et al, Elbes et al and Zhou et al demonstrate the utility of these approaches; however, there are some caveats to note. Firstly, these models may not be able to correctly classify ambiguous cases from images. Using a method where the likelihood is presented versus a definitive category may help in assessing ambiguous cases, where the final decision is made by the practitioner. This is especially important when determining if a patient needs microbiological testing, as missing possible positive cases may lead to delayed testing, potentially leading to poor outcomes for infected patients.³

In addition to diagnostic applications, these models have the potential to assist in patient management planning and treatment decisions. Various commercial versions of ChatGPT (OpenAI) have been tested as diagnostic agents and aids, patient education tools, and in teledermatology applications.³⁴ These general models have been found to perform reasonably well on medical licencing exams, with Mihalache et al demonstrating that ChatGPT-4 was able to correctly answer 88% text-based questions from the United States Medical Licencing Examination practice set.³⁵ Very recently, Zhan et al compared GPT-4o to both infectious disease residents and specialists on a set of panel-developed questions, finding that GPT-4o performed at an equivalent level to specialists, demonstrating its potential as a tool for medical

students and junior residents.³⁶ Fine tuning these existing models with dermatology and medical information will allow for these models to become effective tools for overburdened physicians. These LLMs may be able to provide sufficient and accurate responses for physicians on possible patient diagnoses, suggest appropriate diagnostic tests and even provide treatment plans that the physician can then adjust to the patient's needs.³⁷

While specific clinical application of ChatGPT and other commercial LLMs as physician aids in the management of superficial fungal infections has not been found in our literature search, there is evidence of their utility in other dermatological applications. Liu et al examined the performance of both ChatGPT-4 with vision and Claude 3 Opus in melanoma assessment, with Claude 3 Opus outperforming ChatGPT with an accuracy of 56% and 48%, respectively.³⁸ This type of approach can be adapted to the diagnosis and management of superficial fungal infections and careful integration of these LLMs into clinical practice would benefit both patient and physician.³⁹

Due to their unique capabilities, ChatGPT, as well as other commercial LLMs, have great potential in teledermatology applications. In 2024, Shapiro and Lyakovitsky demonstrated the utility of a teledermatology chatbot (“Dr. DermBot”) which uses ChatGPT-4 as a backbone.⁴⁰ This chatbot was designed to take the patient’s medical history, examine and classify provided clinical images, and then finally provide recommendations. The model was able to accurately diagnose alopecia areata in a patient test case.⁴⁰ These types of models, as well as those described earlier by Han et al, Zhou et al, Abdulhadi et al, Elbes et al and Kim et al, would fill essential care gaps in rural and remote areas, but would also provide patients with continued access to care in situations where patient-physician contact is not possible, such as in the COVID-19 pandemic.^{28,29,31–33,37,41}

Additionally, these LLMs can act as patient education tools, informing patients about what types of outcome they can expect with their disorder or infections, as well as providing information to patients regarding appropriate care strategies for their condition.^{42–44}

Altogether, AI technologies and LLMs specifically have the capacity to greatly increase patient access to dermatological care, specifically supporting patients in underserved regions. However, appropriate implementation and use are essential.

What We Know and Where We are Going

The current landscape of AI technology is extremely dynamic. A search of clinical trial registries reveals numerous studies at different stages of completion addressing a range of clinical applications and workflows, including but definitely not limited to a trial of an AI decision support system to reduce unnecessary antimicrobial prescriptions in children (NCT0687625),⁴⁵ a study of AI in supporting skin cancer management (NCT04040114),⁴⁶ AI implementation in teledermoscopy (NCT05033678)⁴⁷ and a trial for clinical workflow optimization with AI in dermatological conditions (NCT06263413).⁴⁸

Currently, there are only a limited number of studies of specifically exploring the use of LLMs in the management of superficial fungal infections, with published studies assessing diagnostic accuracy or demonstrating feasibility.^{13,14,26–28} Larger studies to specifically assess model accuracy, along with much larger sets of training data, will be need to progress development in this particular niche.

A range of AI tools are already being used clinically, namely AI medical scribe software such as, but not limited to, DeepScribe (www.deepscribe.ai), Lyrebird (lyrebirdhealth.com), Tali (tali.ai) and Scribeberry (scribeberry.com), assisting physicians with documenting patient interactions and administrative tasks. These clinically available AI tools, as well as those in development have the potential to significantly lower physician workload and increase diagnostic accuracy. Additionally, teledermatology-based AI tools would improve access to dermatological care, particularly in regions where adequate medical resources may not be available.

Even though these technologies possess great potential, there are significant challenges that will need to be addressed for this potential to be realized. Current models can experience hallucinations, particularly when the model has not seen similar information during the training process.^{21,22} Additionally, bias may be introduced into these models during the training phase through skewed training data, both of which have the potential to lead to patient harm. Another issue is that there is a regulatory “grey zone” surrounding the use of these models.²² Commercial generic LLMs, such as ChatGPT, which may be currently used by physicians, have not been specifically developed for medical use and would

not have undergone the necessary regulatory processes, raising the concern for potential patient harm and legal implications around responsibility for medical errors.^{22,23,25}

At the current rate of LLM development, what is state-of-the-art today may be obsolete within a year's time. Current estimates are that performance is doubling between every 3 to 14 months, with the majority of estimates at approximately 7 to 8 months.^{49–51} This rapid development holds great promise for improving performance in dermatological applications; however, there is a great risk that this extremely fast development will outpace not only the ability of physicians to train on these systems, but also the ability of governments to establish requirements and regulations around their safe use. The potential benefits of this technology can outweigh the risks, but only if we choose to as a society.

Disclosure

The authors have no conflict of interest to declare. The authors received no funding for this work.

References

1. Finch J, Erin W. Toenail onychomycosis: current and future treatment options. *Dermatol Ther.* 2007;20(1):31–46. doi:10.1111/j.1529-8019.2007.00109.x
2. Khan SS, Hay R, Saunte DML. An international survey of recalcitrant and recurrent tinea of the glabrous skin—A potential indicator of antifungal resistance. *J Eur Acad Dermatol Venereol.* 2025;39(6):1185–1191. doi:10.1111/jdv.20146
3. Gupta AK, Tu LQ. Onychomycosis therapies: strategies to improve efficacy. *Dermatol Clin.* 2006;24(3):381–386. doi:10.1016/j.det.2006.03.009
4. Tamimi P, Fattahi M, Ghaderi A, et al. Terbinafine-resistant *T. indotineae* due to F397L/L393S or F397L/L393F mutation among corticoid-related tinea incognita patients. *J Dtsch Dermatol Ges.* 2024;22(7):922–934. doi:10.1111/ddg.15440
5. Briganti G, Le Moine O. Artificial intelligence in medicine: today and tomorrow. *Front Med.* 2020;7(February):1–6. doi:10.3389/fmed.2020.00027
6. Seshadri DR, Bittel B, Browsky D, et al. Accuracy of apple watch for detection of atrial fibrillation. *Circulation.* 2020;141(8):702–703. doi:10.1161/CIRCULATIONAHA.119.044126
7. Turakhia MP, Desai M, Hedlin H, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: the apple heart study. *Am Heart J.* 2019;207:66–75. doi:10.1016/j.ahj.2018.09.002
8. Goodfellow I, Bengio Y, Courville A. *Deep Learning.* MIT Press; 2016; doi:10.1007/978-981-13-9113-2_16
9. Charniak E. *Introduction to Deep Learning.* The MIT Press; 2018.
10. Qiu Y. The impact of LLM hallucinations on motor skill learning: a case study in badminton. *IEEE Access.* 2024;139669–139682. doi:10.1109/ACCESS.2024.3444783.
11. Sathe N, Deodhe V, Sharma Y, Shinde A. A comprehensive review of ai in healthcare: exploring neural networks in medical imaging, LLM-based interactive response systems, NLP-based EHR systems, ethics, and beyond. In: *2023 International Conference on Advanced Computing & Communication Technologies (ICACCTech).* IEEE; 2023:633–640. doi:10.1109/ICACCTech61146.2023.00108.
12. Kim TW. Application of artificial intelligence chatbots, including ChatGPT, in education, scholarly work, programming, and content generation and its prospects: a narrative review. *J Educ Eval Health Prof.* 2023;20:38. doi:10.3352/jeehp.2023.20.38
13. Koo T, Kim MH, Jue MS. Automated detection of superficial fungal infections from microscopic images through a regional convolutional neural network. *PLoS One.* 2021;16(8 August):1–11. doi:10.1371/journal.pone.0256290
14. Zielinski B, Sroka-Oleksiak A, Rymarczyk D, Piekarczyk A, Brzychczy-Woch M. Deep learning approach to describe and classify fungi microscopic images. *PLoS One.* 2020;15(6 June):1–16. doi:10.1371/journal.pone.0234806
15. Weidinger L, Mellor J, Rauh M, et al. Ethical and social risks of harm from language models. *ArXiv.* 2021. 1–64.
16. Luo S, Yu A, Xie Z, et al. emGene: an embodied LLM NGS sequencer for real-time precision diagnostics. *Integr Circuits Syst.* 2025;2(2):67–80. doi:10.23919/ICS.2025.3552542
17. Han T, Adams LC, Papaioannou JM, et al. MedAlpaca – an open-source collection of medical conversational AI models and training data. *ArXiv.* 2023. 1–9.
18. Wu C, Lin W, Zhang X, Zhang Y, Xie W, Wang Y. PMC-LLaMA: toward building open-source language models for medicine. *J Am Med Inform Assoc.* 2024;31(9):1833–1843. doi:10.1093/jamia/ocae045
19. Li J, Deng Y, Sun Q, et al. Benchmarking large language models in evidence-based medicine. *IEEE J Biomed Health Inform.* 2024;1–14. doi:10.1109/JBHI.2024.3483816.
20. Smith KP, Kirby JE. Image analysis and artificial intelligence in infectious disease diagnostics. *Clin Microbiol Infect.* 2020;26(10):1318–1323. doi:10.1016/j.cmi.2020.03.012
21. Schwartz IS, Link KE, Daneshjou R, Cortés-Penfield N. Black box warning: large language models and the future of infectious diseases consultation. *Clin Infect Dis.* 2024;78(4):860–866. doi:10.1093/cid/ciad633
22. Sarantopoulos A, Mastori Kourmpanti C, Yokarasa AL, et al. Artificial intelligence in infectious disease clinical practice: an overview of gaps, opportunities, and limitations. *Trop Med Infect Dis.* 2024;9(10):228. doi:10.3390/tropicalmed9100228
23. Siafakas N, Vasarmidi E. Risks of Artificial Intelligence (AI) in medicine. *Pneumon.* 2024;37(3):1–5. doi:10.18332/pne/191736
24. OpenAI. Memory FAQ | openAI Help Center. Available from: https://help.openai.com/en/articles/8590148-memory-faq#h_da6d3fae16. Accessed June 4, 2025.
25. Goh E, Gallo R, Strong E, et al. Large language model influence on management reasoning: a randomized controlled trial. *medRxiv.* 2024;7(10):1–12. doi:10.1101/2024.08.05.24311485
26. Jansen P, Creosteanu A, Matyas V, et al. Deep learning assisted diagnosis of onychomycosis on whole-slide images. *J Fungi.* 2022;8(9):912. doi:10.3390/jof8090912

27. Decroos F, Springenberg S, Lang T, et al. A deep learning approach for histopathological diagnosis of onychomycosis: not inferior to analogue diagnosis by histopathologists. *Acta Derm Venereol.* 2021;101(8):1–8. doi:10.2340/00015555-3893
28. Han SS, Park GH, Lim W, et al. Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: automatic construction of onychomycosis datasets by region-based convolutional deep neural network. *PLoS One.* 2018;13(1):1–14. doi:10.1371/journal.pone.0191493
29. Kim YJ, Han SS, Yang HJ, Chang SE. Prospective, comparative evaluation of a deep neural network and dermoscopy in the diagnosis of onychomycosis. *PLoS One.* 2020;15(6):e0234334. doi:10.1371/journal.pone.0234334
30. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017.* 2016:2261–2269. doi:10.1109/CVPR.2017.243.
31. Abdulhadi J, Al-Dujaili A, Humaidi AJ, Fadhel MAR. Human nail diseases classification based on transfer learning. *ICIC Express Letters.* 2021;15(12):1271–1282. doi:10.24507/icicel.15.12.1271
32. Elbes M, AlZu'bi S, Kanan T, Mughaid A, Abushanab S. Big dermatological data service for precise and immediate diagnosis by utilizing pre-trained learning models. *Cluster Comput.* 2024;27(5):6931–6951. doi:10.1007/s10586-024-04331-8
33. Zhou J, He X, Sun L, et al. SkinGPT-4: an interactive dermatology diagnostic system with visual large language model. *ArXiv.* 2023:1–12.
34. Goktas P, Grzybowski A. Assessing the impact of ChatGPT in dermatology: a comprehensive rapid review. *J Clin Med.* 2024;13(19):5909. doi:10.3390/jcm13195909
35. Mihalache A, Huang RS, Popovic MM, Muni RH. ChatGPT-4: an assessment of an upgraded artificial intelligence chatbot in the United States medical licensing examination. *Med Teach.* 2024;46(3):366–372. doi:10.1080/0142159X.2023.2249588
36. Zhan L, Dang X, Xie Z, et al. Evaluating GPT-4o in infectious disease diagnostics and management: a comparative study with residents and specialists on accuracy, completeness, and clinical support potential. *Digit Health.* 2025;11:1–9. doi:10.1177/20552076251355797
37. Ahmed SK, Hussein S, Aziz TA, Chakraborty S, Islam MR, Dhama K. The power of ChatGPT in revolutionizing rural healthcare delivery. *Health Sci Rep.* 2023;6(11):2–5. doi:10.1002/hsr2.1684
38. Liu X, Duan C, Kim MK, et al. Claude 3 opus and ChatGPT with GPT-4 in dermoscopic image analysis for melanoma diagnosis: comparative performance analysis. *JMIR Med Inform.* 2024;12:1–4. doi:10.2196/59273
39. Gisselbaek M, Berger-Estilita J, Devos A, Ingrassia PL, Dieckmann P, Saxena S. Bridging the gap between scientists and clinicians: addressing collaboration challenges in clinical AI integration. *BMC Anesthesiol.* 2025;25(1):269. doi:10.1186/s12871-025-03130-x
40. Shapiro J, Lyakhovitsky A. Revolutionizing teledermatology: exploring the integration of artificial intelligence, including generative pre-trained transformer chatbots for artificial intelligence-driven anamnesis, diagnosis, and treatment plans. *Clin Dermatol.* 2024;42(5):492–497. doi:10.1016/j.clindermatol.2024.06.020
41. Ruggiero A, Martora F, Fabbrocini G, et al. The role of teledermatology during the COVID-19 pandemic: a narrative review. *Clin Cosmet Invest Dermatol.* 2022;15(December):2785–2793. doi:10.2147/CCID.S377029
42. Lakdawala N, Channa L, Gronbeck C, et al. Assessing the accuracy and comprehensiveness of ChatGPT in offering clinical guidance for atopic dermatitis and acne vulgaris. *JMIR Dermatol.* 2023;6(1):e50409. doi:10.2196/50409
43. Ferreira AL, Chu B, Grant-Kels JM, Ogunleye T, Lipoff JB. Evaluation of ChatGPT dermatology responses to common patient queries. *JMIR Dermatol.* 2023;6(1):1–3. doi:10.2196/49280
44. Mondal H, Mondal S, Podder I. Using ChatGPT for writing articles for patients' education for dermatological diseases: a pilot study. *Indian Dermatol Online J.* 2023;14(4):482. doi:10.4103/idoj.idoj_72_23
45. Shope T. Artificial intelligence diagnostic decision support to reduce antimicrobial prescriptions in young children with colds (IMAGE). Available from: <https://clinicaltrials.gov/study/NCT06876259>. Accessed August 13, 2025.
46. Melanoma and Skin Cancer Trials Limited. Improving skin cancer management with artificial intelligence (04.17 SMARTI) (SMARTI). Available from: <https://clinicaltrials.gov/study/NCT04040114>. Accessed August 13, 2025.
47. Region Skane. Implementation of teledermoscopy and artificial intelligence. Available from: <https://clinicaltrials.gov/study/NCT05033678>. Accessed August 13, 2025.
48. AI Labs Group S.L. Clinical workflow optimization using artificial intelligence for dermatological conditions (IDEI_2023). Available from: <https://clinicaltrials.gov/study/NCT06263413>. Accessed August 13, 2025.
49. Ho A, Besiroglu T, Erdil E, et al. Algorithmic progress in language models. *ArXiv.* 2024:1–31.
50. Polo FM, Somerstep S, Choshen L, Sun Y, Yurochkin M. Sloth: scaling laws for LLM skills to predict multi-benchmark performance across families. *ArXiv.* 2024.
51. Xiao C, Cai J, Zhao W, et al. Densing Law of LLMs. *ArXiv.* 2024:1–16.

Clinical, Cosmetic and Investigational Dermatology

Publish your work in this journal

Clinical, Cosmetic and Investigational Dermatology is an international, peer-reviewed, open access, online journal that focuses on the latest clinical and experimental research in all aspects of skin disease and cosmetic interventions. This journal is indexed on CAS. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/clinical-cosmetic-and-investigational-dermatology-journal>

Dovepress
Taylor & Francis Group