# How can large language models assist your research?

Collin Holgate

Created with DALL-E 3

OPINION

GUEST ESSAY

# One Year In and ChatGPT Already Has Us Doing Its Bidding

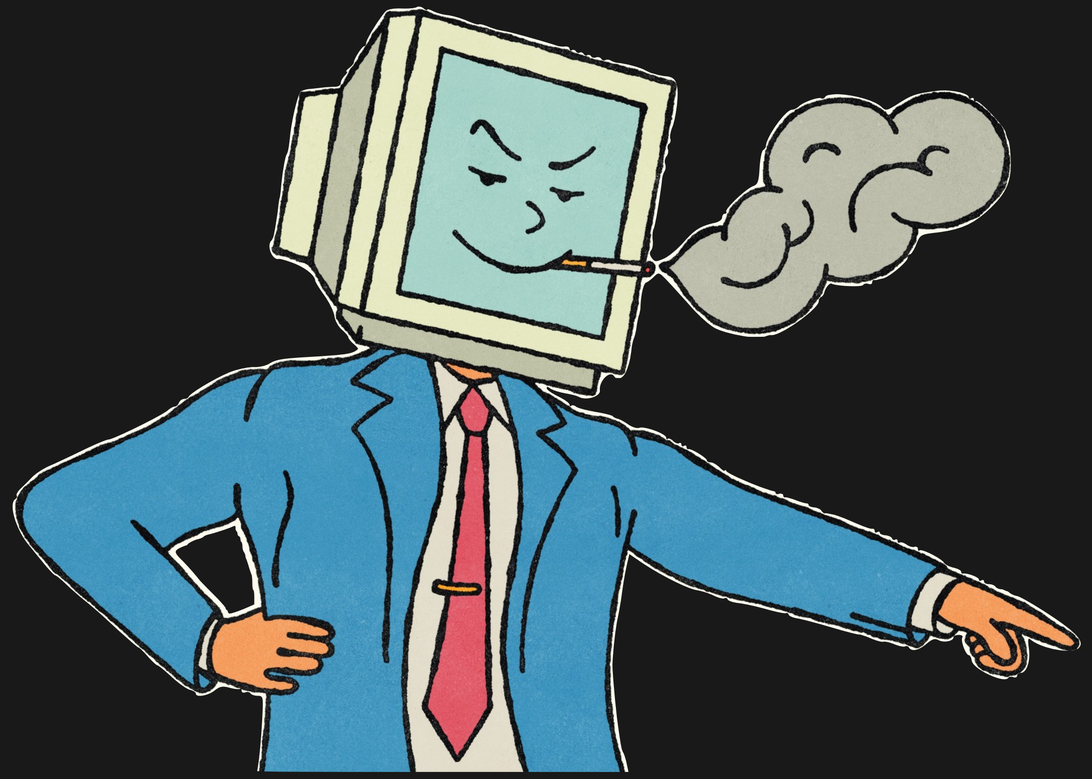
Emma Grillo, NY Times

TECHNOLOGY

# DOES SAM ALTMAN KNOW WHAT HE'S CREATING?

The OpenAI CEO's ambitious, ingenious, terrifying quest to create a new form of intelligence

Ross Anderson, The Atlantic



Zak Tebbal, NY Times

# How do LLMs work?

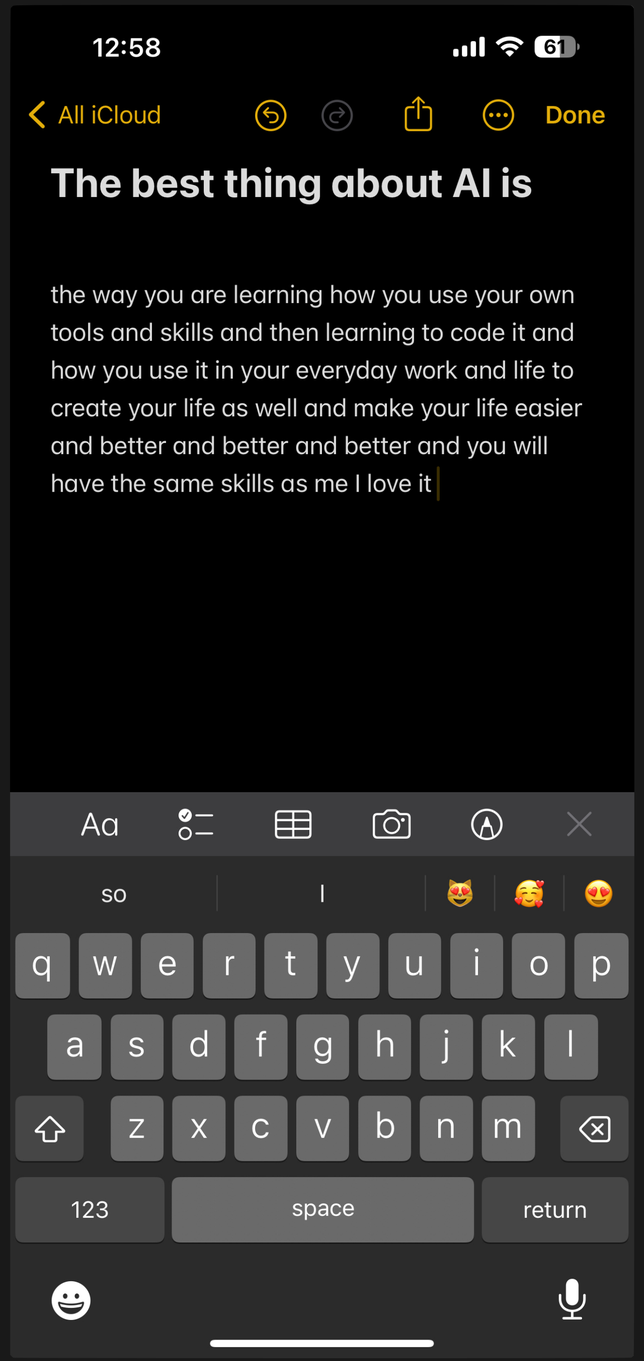
# Large language models generate text one word (token) at a time

Effectively, the model asks:

“For the combination of words that came before, what is likely to come next?”

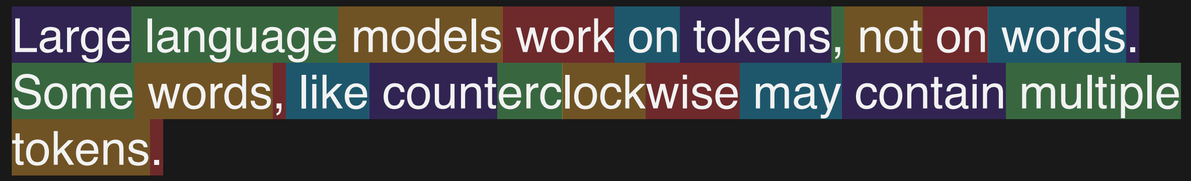
A simple analogy would be continuously mashing the auto-suggestion keyboard on a smartphone!

But, the models play some clever tricks and do some clever things to produce higher-quality output.



# Clever tricks allow the models to be smarter and more efficient

The models break down words into component pieces called tokens to improve storage and processing. More complicated words can be described as components!




The models don’t just pick the most likely token—they pick from a list of most-likely tokens to include some “randomness”. This effectively gives them “creativity”.

The best thing about AI is its ability to…

| | |
|---|---|
| learn | 4.5% |
| predict | 3.5% |
| make | 3.2% |
| understand | 3.1% |
| do | 2.9% |

Example from “What is ChatGPT Doing…and Why Does it Work? By Stephen Wolfram

# Different models have distinct training sets and “philosophies” baked into them, so they behave uniquely!

*The best thing about AI is…*

its ability to learn and improve to better serve humanity. With the right **oversight and ethical guidelines**, AI has immense potential to make positive contributions in areas like healthcare, education, transportation, and more. However, we must ensure AI develops **safely** and for the benefit of all people. If **guided properly**, AI can be a powerful tool to **create a more just, prosperous and sustainable world**.

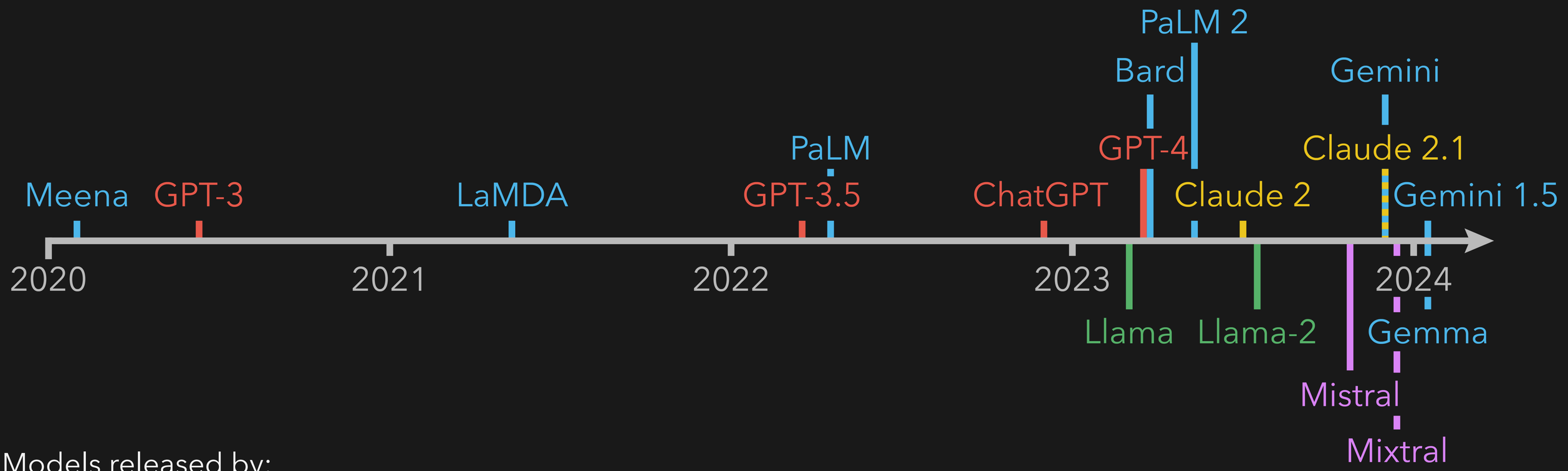
Predicted with Claude 2.1

its ability to **revolutionize countless aspects of our lives** by processing data at speeds and scales beyond human capability. It automates routine tasks and personalizes experiences, from education to shopping, **freeing humans** to tackle more complex challenges. As AI technology advances, it promises to drive progress in fields like healthcare, environmental sustainability, and public safety, **showcasing its vast potential for positive impact**.

Predicted with GPT-4

# What tools are available?

# LLMs are evolving quickly and drastically



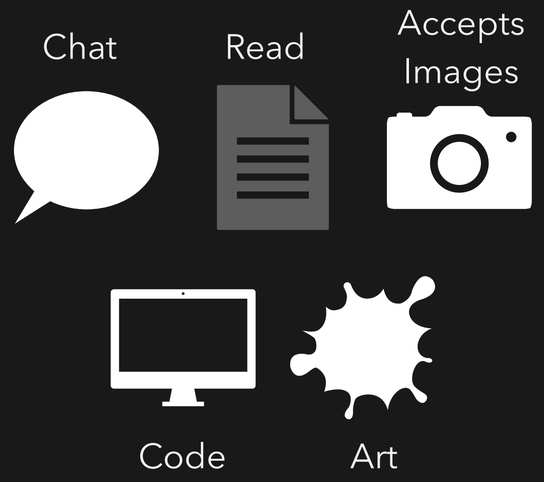

Models released by:

- Google
- OpenAI
- Meta
- Anthropic
- Mistral

# State-of-the-art *closed* LLMs

### OpenAI

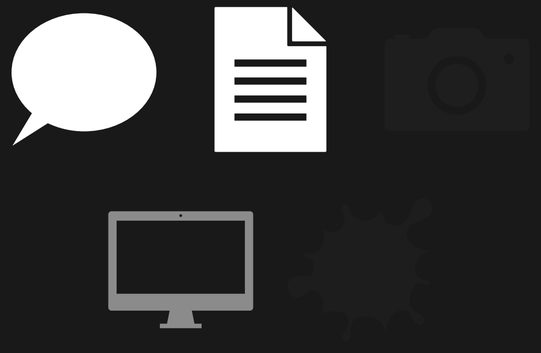
- Accessible through ChatGPT
- Paid tier at $20/mo




The most used

### Anthropic

- Accessible through Claude
- Paid tier at $20/mo



Best free tier

### Google (Alphabet)

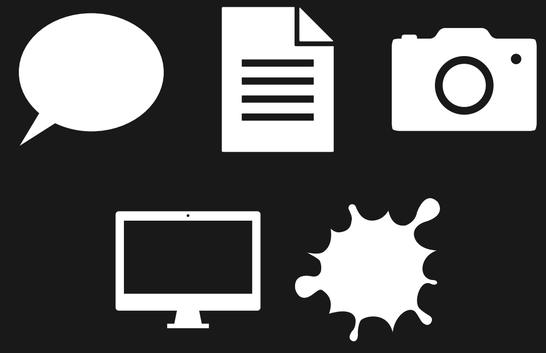
- Accessible through Gemini
- Paid tier at $20/mo
- Includes fact-checking feature via Google

Great connectivity

### Combination tools

Perplexity.ai

- Access to GPT-4 and Claude 2.1, etc.
- Web connected
- Paid tier at $20/mo



Flexibility

# But these closed models require you to *upload* your information

To some people, this won't be a big deal at all. Your data will likely be lost in a sea of others, but it is worth pointing out.

For the concerned or for the enthusiast, some open-source models have been published that you can run *locally* on your own hardware.*

- Ollama is a great tool for managing, running, and implementing these models.

**Generally these require a **great** computer.*

The quality of Large Language Models generally correlates to the number of parameters (knobs) they were trained using.

| Model Size | Ram Required |
|---|---|
| ~2 B | 3 GB |
| ~7 B | 8 GB |
| ~13 B | 16 GB |
| ~70 B | Uff-da |

# Available open models

### Meta

- Publishes Llama 2
- Available in 13B and 70B configurations

### Mistral

- Publishes Mistral (7B) and Mixtral (8x7B)
- Mixtral uses “Mixture of Experts” approach

### Google (Alphabet)

- Publishes Gemma
- Available in 2B and 7B configurations
- 7 days old!

### Key benefits

- No data uploaded to the cloud
- Open weights:
  - Allow for fine-tuning to achieve certain goals or enhance performance
  - Remove annoying censors

These models require you to do a lot of leg work (e.g., extract your own PDFs)

# How can we use LLMs as a tool?

# We can use LLMs to help us process text

Today we'll cover four main applications I have used these models for (and hopefully have time for others to share theirs!)

1. Text summarization or extraction
2. Writing aids
3. Correction of OCR text
4. Converting SEM-EDS data from .docx or .pdf into a .csv file

*LLM's can be great at helping you code too! I haven't played with this aspect much but I have used it to help debug code.*

# LLMs can process large bodies of text for summarization

Journal of the European Ceramic Society 41 (2021) 1984–1994

Contents lists available at ScienceDirect

Journal of the European Ceramic Society

journal homepage: www.elsevier.com/locate/jeurceramsoc

Original article

Dissolution and diffusion kinetics of yttria-stabilized zirconia into molten silicates

Collin S. Holgate [a,*], Gareth G.E. Seward [b], Andrew R. Ericks [a], David L. Poerschke [c], Carlos G. Levi [a]

[a] Materials Department, University of California, Santa Barbara, CA 93106-5050, United States
[b] Department of Earth Science, University of California, Santa Barbara, CA 93106-9630, United States
[c] Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN, 55455-0132, United States

ARTICLE INFO

*Keywords:*
Yttria-stabilized zirconia
Molten silicate attack
CMAS
Diffusion
Dissolution

ABSTRACT

The degradation of thermal barrier coatings (TBCs) by molten silicates (CMAS) represents a fundamental barrier to progress in gas turbine technology, requiring a mechanistic understanding of the problem to guide the development of improved coatings. This article investigates the dissolution of yttria-stabilized zirconia (7YSZ and 20YSZ) into two model silicate melts at 1300–1400 °C. The approach involves the 1D dissolution of YSZ into a semi-infinite melt, characterizing the dissolution rates of YSZ and the diffusion rates of $Zr^{4+}$ and $Y^{3+}$ therein. The assessed kinetics of YSZ dissolution and diffusion were then applied to modeling the same phenomena on TBC-relevant length scales. These findings provide fundamental insight into (*i*) the dissolution mechanism of YSZ, (*ii*) the subsequent reprecipitation upon saturation, (*iii*) the quantitative effects of temperature and melt composition on the dissolution and diffusion kinetics, and (*iv*) how the measured kinetics manifests on the scale of flow channels present in TBCs.

## 1. Introduction

Thermal barrier coatings (TBCs) based on $ZrO_2$–7 ± 1 wt.% $YO_{1.5}$ (7YSZ) are increasingly challenged for service at the higher temperatures required to meet the efficiency targets in advanced gas turbines [1,2]. A fundamental barrier to progress is the degradation of TBCs by molten silicate deposits, resulting from the ingestion of siliceous debris with the intake air, especially in aircraft engines [3,4]. The silicate melts, generically known as CMAS for their major oxide constituents (Ca, Mg, Al, Si) penetrate the compliance-enabling porosity of the TBC, stiffening the coating upon solidification and generating enough strain energy upon cooling to drive cracking and spallation [4–6].

The leading mitigation strategy for CMAS damage is based on reactive crystallization, wherein the infiltrating melt dissolves the thermal barrier oxide (TBO) and yields crystalline reaction products in sufficient amounts to fill the porosity and arrest melt penetration close to the surface [6,7]. Preeminent TBO examples are rare-earth (RE) zirconates that react with melts over a wide range of Ca:Si ratios (~0.21–0.73) [8] to form oxyapatites based on $Ca_2RE_8Si_6O_{26}$ [7]. In contrast, 7YSZ does not undergo reactive crystallization with the higher Ca:Si melts, although it does form $ZrSiO_4$ (zircon) with the more acidic melts (low Ca:Si) [8]. However, the latter reaction is too sluggish to be competitive with the melt penetration rate, rendering 7YSZ ineffective in CMAS mitigation [6,9]. Instead, dissolved YSZ eventually reprecipitates forming a Y-depleted $ZrO_2$—this phase does not incorporate substantial components of the melt or generate a sufficient volume to block the flow. While the phenomenology of these interactions is established, there is a paucity of fundamental understanding of the mechanisms and the associated kinetics to enable the development of comprehensive models to guide coating design [6]. That need motivates the present study.

The envisaged steps in the mitigation mechanism are: (*i*) dissolution of the TBO upon contact with the melt; (*ii*) build-up of the reactive cation (*s*) at the ensuing boundary layer, mediated by the balance between the dissolution rate and the diffusional dissipation into the bulk melt; (*iii*) nucleation of the desirable crystalline phase(s) incorporating the key components of the TBO and the melt; and (*iv*) growth of the crystalline products to fill the flow channels and stop penetration. For optimum effectiveness the rates of these steps should all be as fast as possible, so the times allowed for melt penetration are minimized. In those cases,

* Corresponding author.
*E-mail address:* holgate@ucsb.edu (C.S. Holgate).

https://doi.org/10.1016/j.jeurceramsoc.2020.10.056
Received 19 May 2020; Received in revised form 23 October 2020; Accepted 24 October 2020
Available online 28 October 2020


*Can you summarize this document? What are the key scientific findings?*

This paper investigates the dissolution and diffusion kinetics of yttria-stabilized zirconia (YSZ) into molten silicate melts. The key findings are:

- There is an initial transient period where the YSZ dissolution rate is controlled by both the interface dissolution rate and diffusion of dissolved ions away from the interface. During this period, the concentrations of dissolved YSZ components at the interface increase over time.
- In polycrystalline YSZ, preferential dissolution occurs along grain boundaries, leading to a "mushy" reaction zone. Y-depleted zirconia reprecipitates epitaxially around undissolved cores, forming core-shell structures.

*Truncated…*

The model is pulling information from beyond just the abstract and conclusions!



*Calculated with Claude 2 (free)*

# LLMs can process large bodies of text or for analysis

Journal of the European Ceramic Society 41 (2021) 1984–1994

Contents lists available at ScienceDirect

Journal of the European Ceramic Society

journal homepage: www.elsevier.com/locate/jeurceramsoc


Original article

## Dissolution and diffusion kinetics of yttria-stabilized zirconia into molten silicates

Collin S. Holgate [a,*], Gareth G.E. Seward [b], Andrew R. Ericks [a], David L. Poerschke [c], Carlos G. Levi [a]

[a] Materials Department, University of California, Santa Barbara, CA 93106-5050, United States
[b] Department of Earth Science, University of California, Santa Barbara, CA 93106-9630, United States
[c] Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN, 55455-0132, United States

ARTICLE INFO

*Keywords:*
Yttria-stabilized zirconia
Molten silicate attack
CMAS
Diffusion
Dissolution

ABSTRACT

The degradation of thermal barrier coatings (TBCs) by molten silicates (CMAS) represents a fundamental barrier to progress in gas turbine technology, requiring a mechanistic understanding of the problem to guide the development of improved coatings. This article investigates the dissolution of yttria-stabilized zirconia (7YSZ and 20YSZ) into two model silicate melts at 1300–1400 °C. The approach involves the 1D dissolution of YSZ into a semi-infinite melt, characterizing the dissolution rates of YSZ and the diffusion rates of $Zr^{4+}$ and $Y^{3+}$ therein. The assessed kinetics of YSZ dissolution and diffusion were then applied to modeling the same phenomena on TBC-relevant length scales. These findings provide fundamental insight into (*i*) the dissolution mechanism of YSZ, (*ii*) the subsequent reprecipitation upon saturation, (*iii*) the quantitative effects of temperature and melt composition on the dissolution and diffusion kinetics, and (*iv*) how the measured kinetics manifests on the scale of flow channels present in TBCs.

### 1. Introduction

Thermal barrier coatings (TBCs) based on $ZrO_2$–7 ± 1 wt.% $YO_{1.5}$ (7YSZ) are increasingly challenged for service at the higher temperatures required to meet the efficiency targets in advanced gas turbines [1,2]. A fundamental barrier to progress is the degradation of TBCs by molten silicate deposits, resulting from the ingestion of siliceous debris with the intake air, especially in aircraft engines [3,4]. The silicate melts, generically known as CMAS for their major oxide constituents (Ca, Mg, Al, Si) penetrate the compliance-enabling porosity of the TBC, stiffening the coating upon solidification and generating enough strain energy upon cooling to drive cracking and spallation [4–6].

The leading mitigation strategy for CMAS damage is based on reactive crystallization, wherein the infiltrating melt dissolves the thermal barrier oxide (TBO) and yields crystalline reaction products in sufficient amounts to fill the porosity and arrest melt penetration close to the surface [6,7]. Preeminent TBO examples are rare-earth (RE) zirconates that react with melts over a wide range of Ca:Si ratios (~0.21–0.73) [8] to form oxyapatites based on $Ca_2RE_8Si_6O_{26}$ [7]. In contrast, 7YSZ does not undergo reactive crystallization with the higher Ca:Si melts, although it does form $ZrSiO_4$ (zircon) with the more acidic melts (low Ca:Si) [8]. However, the latter reaction is too sluggish to be competitive with the melt penetration rate, rendering 7YSZ ineffective in CMAS mitigation [6,9]. Instead, dissolved YSZ eventually reprecipitates forming a Y-depleted $ZrO_2$—this phase does not incorporate substantial components of the melt or generate a sufficient volume to block the flow. While the phenomenology of these interactions is established, there is a paucity of fundamental understanding of the mechanisms and the associated kinetics to enable the development of comprehensive models to guide coating design [6]. That need motivates the present study.

The envisaged steps in the mitigation mechanism are: (*i*) dissolution of the TBO upon contact with the melt; (*ii*) build-up of the reactive cation (*s*) at the ensuing boundary layer, mediated by the balance between the dissolution rate and the diffusional dissipation into the bulk melt; (*iii*) nucleation of the desirable crystalline phase(s) incorporating the key components of the TBO and the melt; and (*iv*) growth of the crystalline products to fill the flow channels and stop penetration. For optimum effectiveness the rates of these steps should all be as fast as possible, so the times allowed for melt penetration are minimized. In those cases,

* Corresponding author.
*E-mail address:* holgate@ucsb.edu (C.S. Holgate).

https://doi.org/10.1016/j.jeurceramsoc.2020.10.056
Received 19 May 2020; Received in revised form 23 October 2020; Accepted 24 October 2020
Available online 28 October 2020


*How did they handle any issues with convection?*

The paper discusses some steps taken to minimize convection effects on the measured diffusion profiles:

- The experimental setup constrained the dissolution area to the top of the YSZ coupon, reducing melt density differences that could drive convection.
- Multiple line scans were taken across each diffusion profile to average out any variations from convection.

*Truncated…*



*Calculated with Claude 2 (free)*

# LLMs can process large bodies of text or for analysis

Journal of the European Ceramic Society 41 (2021) 1984–1994

Contents lists available at ScienceDirect

Journal of the European Ceramic Society

journal homepage: www.elsevier.com/locate/jeurceramsoc


Original article

## Dissolution and diffusion kinetics of yttria-stabilized zirconia into molten silicates

Collin S. Holgate [a,*], Gareth G.E. Seward [b], Andrew R. Ericks [a], David L. Poerschke [c], Carlos G. Levi [a]

[a] Materials Department, University of California, Santa Barbara, CA 93106-5050, United States
[b] Department of Earth Science, University of California, Santa Barbara, CA 93106-9630, United States
[c] Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN, 55455-0132, United States

ARTICLE INFO

*Keywords:*
Yttria-stabilized zirconia
Molten silicate attack
CMAS
Diffusion
Dissolution

ABSTRACT

The degradation of thermal barrier coatings (TBCs) by molten silicates (CMAS) represents a fundamental barrier to progress in gas turbine technology, requiring a mechanistic understanding of the problem to guide the development of improved coatings. This article investigates the dissolution of yttria-stabilized zirconia (7YSZ and 20YSZ) into two model silicate melts at 1300–1400 °C. The approach involves the 1D dissolution of YSZ into a semi-infinite melt, characterizing the dissolution rates of YSZ and the diffusion rates of $Zr^{4+}$ and $Y^{3+}$ therein. The assessed kinetics of YSZ dissolution and diffusion were then applied to modeling the same phenomena on TBC-relevant length scales. These findings provide fundamental insight into (*i*) the dissolution mechanism of YSZ, (*ii*) the subsequent reprecipitation upon saturation, (*iii*) the quantitative effects of temperature and melt composition on the dissolution and diffusion kinetics, and (*iv*) how the measured kinetics manifests on the scale of flow channels present in TBCs.

### 1. Introduction

Thermal barrier coatings (TBCs) based on $ZrO_2$–7 ± 1 wt.% $YO_{1.5}$ (7YSZ) are increasingly challenged for service at the higher temperatures required to meet the efficiency targets in advanced gas turbines [1,2]. A fundamental barrier to progress is the degradation of TBCs by molten silicate deposits, resulting from the ingestion of siliceous debris with the intake air, especially in aircraft engines [3,4]. The silicate melts, generically known as CMAS for their major oxide constituents (Ca, Mg, Al, Si) penetrate the compliance-enabling porosity of the TBC, stiffening the coating upon solidification and generating enough strain energy upon cooling to drive cracking and spallation [4–6].

The leading mitigation strategy for CMAS damage is based on reactive crystallization, wherein the infiltrating melt dissolves the thermal barrier oxide (TBO) and yields crystalline reaction products in sufficient amounts to fill the porosity and arrest melt penetration close to the surface [6,7]. Preeminent TBO examples are rare-earth (RE) zirconates that react with melts over a wide range of Ca:Si ratios (~0.21–0.73) [8] to form oxyapatites based on $Ca_2RE_8Si_6O_{26}$ [7]. In contrast, 7YSZ does not undergo reactive crystallization with the higher Ca:Si melts, although it does form $ZrSiO_4$ (zircon) with the more acidic melts (low Ca:Si) [8]. However, the latter reaction is too sluggish to be competitive with the melt penetration rate, rendering 7YSZ ineffective in CMAS mitigation [6,9]. Instead, dissolved YSZ eventually reprecipitates forming a Y-depleted $ZrO_2$—this phase does not incorporate substantial components of the melt or generate a sufficient volume to block the flow. While the phenomenology of these interactions is established, there is a paucity of fundamental understanding of the mechanisms and the associated kinetics to enable the development of comprehensive models to guide coating design [6]. That need motivates the present study.

The envisaged steps in the mitigation mechanism are: (*i*) dissolution of the TBO upon contact with the melt; (*ii*) build-up of the reactive cation (*s*) at the ensuing boundary layer, mediated by the balance between the dissolution rate and the diffusional dissipation into the bulk melt; (*iii*) nucleation of the desirable crystalline phase(s) incorporating the key components of the TBO and the melt; and (*iv*) growth of the crystalline products to fill the flow channels and stop penetration. For optimum effectiveness the rates of these steps should all be as fast as possible, so the times allowed for melt penetration are minimized. In those cases,

* Corresponding author.
*E-mail address:* holgate@ucsb.edu (C.S. Holgate).

https://doi.org/10.1016/j.jeurceramsoc.2020.10.056
Received 19 May 2020; Received in revised form 23 October 2020; Accepted 24 October 2020
Available online 28 October 2020


*What are the flaws of this paper?*

Some potential limitations or flaws of this paper:

- The use of model CMAS and CAS compositions may not fully represent the complex melts that form in real turbine environments.
- 1D diffusion modeling likely oversimplifies the actual geometric conditions in porous TBC coatings.
- Grain boundaries in polycrystalline YSZ add complexity versus the cleaner single crystal results.
- Phase composition changes from sintering in the 7YSZ samples may confound some observations.
- The factors controlling the yttria saturation value are not fully resolved.
- The attempts to extend findings to TBC length scales rely on simple modeling.

*Truncated…*



*Calculated with Claude 2 (free)*

# LLMs are useful tools for writing

They are, in my opinion, especially good at text augmentation:

- Suggesting similar words (thesaurus)
- Identifying features missed by traditional spell-check
  - Repetitiveness (word or ideas)
  - Clarity
- Converting passive voice to active voice easily

# LLMs are useful tools for writing

They are, in my opinion, especially good at text augmentation:

- Suggesting similar words (thesaurus)
- Identifying features missed by traditional spell-check
  - Repetitiveness (word or ideas)
  - Clarity
- Converting passive voice to active voice easily

*I need a word similar to "augmentation" that captures the idea of changing text.*

1. **Revision** - This term suggests a thorough review and subsequent changes to improve the text's accuracy, flow, or quality.
2. **Refinement** - This implies making subtle adjustments to improve the text's elegance or effectiveness.
3. **Enhancement** - Similar to augmentation, enhancement suggests adding features, details, or elements to the text to improve its quality or value.
4. **Transformation** - This implies a significant or fundamental change in the text, possibly altering its original meaning or structure.

*Truncated…*

**You can also use this to match desired tones!**

Optimistic: Refinement Pessimistic: Obfuscation

*Calculated with GPT-4*

# LLMs are useful tools for writing

They are, in my opinion, especially good at text augmentation:

- Suggesting similar words (thesaurus)
- Identifying features missed by traditional spell-check
  - Repetitiveness (word or ideas)
  - Clarity
- Converting passive voice to active voice easily

Good performance here requires good prompting! Some trial-and-error helps a lot.

*The text below is a fragment from a scientific publication in preparation. Please suggest edits to the text for clarity and repetitiveness. Assume any unreferenced material will be elsewhere (i.e., you do not have to say references are missing or abbreviations are undefined). The audience are experts in the area presented.*

# LLMs are useful tools for writing

They are, in my opinion, especially good at text augmentation:

- Suggesting similar words (thesaurus)
- Identifying features missed by traditional spell-check
  - Repetitiveness (word or ideas)
  - Clarity
- Converting passive voice to active voice easily

*Please convert this sentence from passive to active voice:*

*"Thermal barrier coatings (TBCs) based on ZrO2–7 ±1 wt.% $YO_{1.5}$ (7YSZ) are increasingly challenged for service at the higher temperatures required to meet the efficiency targets in advanced gas turbines.*

The higher temperatures required to meet the efficiency targets in advanced gas turbines increasingly challenge thermal barrier coatings (TBCs) based on ZrO2–7 ±1 wt.% YO1.5 (7YSZ) for service.



*Calculated with Claude 2.1*

# LLMs are useful tools for writing

They are, in my opinion, especially good at text augmentation:

- Suggesting similar words (thesaurus)
- Identifying features missed by traditional spell-check
  - Repetitiveness (word or ideas)
  - Clarity
- Converting passive voice to active voice easily

*Please make the text below consistent with Strunk & White Guidelines:*

*"Thermal barrier coatings (TBCs) based on $ZrO_2$–7 ±1 wt.% $YO_{1.5}$ (7YSZ) are increasingly challenged for service at the higher temperatures required to meet the efficiency targets in advanced gas turbines."*

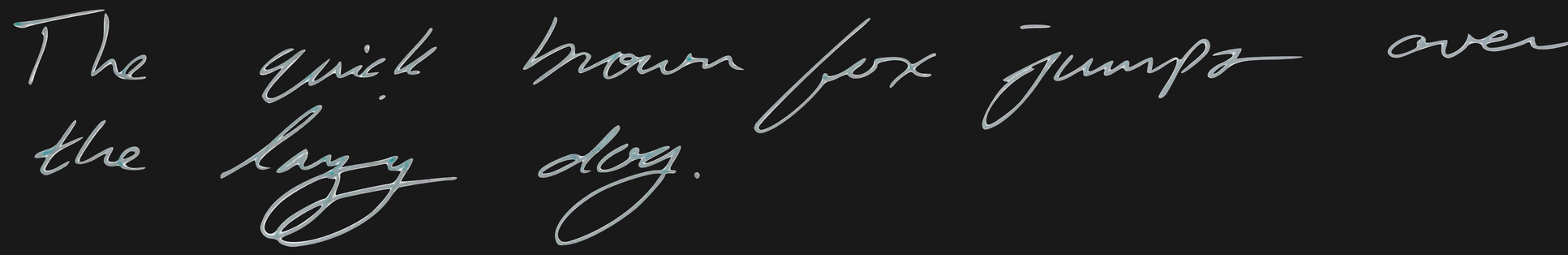
Thermal barrier coatings (TBCs) based on $ZrO_2$–7±1 wt% $YO_{1.5}$ (7YSZ) face increasing challenges in meeting the efficiency targets of advanced gas turbines that require higher service temperatures.

This cut three words and converted to active voice!



*Calculated with GPT-4*

# LLMs can help correct text that was poorly OCR’d




Selecting this text with my iPhone yields:

The the
howr
quick pown ox jump
over
lagy
dog.

*You are going to be given a piece of text that is the result of an OCR conversion from hand written text. The OCR may contain errors that weren't present in the source text. Your job is to fix any errors in this process using your judgement. If you aren't sure what could fill a space, please say so.*

"The quick brown fox jumps over the lazy dog."



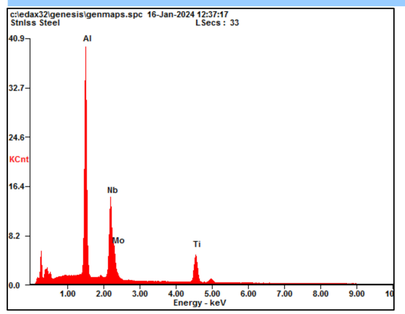
*Calculated with GPT-4*

# My favorite: LLMs can help compile SEM EDS data!

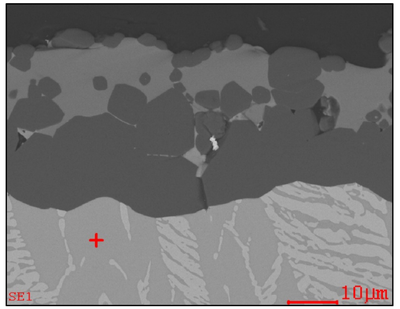
***Microanalysis Report***

***Prepared for:*** ***Company Name Here***

***Prepared by:*** ***Your Name Here*** ***26-Feb-24***




| *Element* | *Wt%* | *At%* |
|---|---|---|
| *AlK* | 30.50 | 52.75 |
| *NbL* | 39.59 | 19.88 |
| *MoL* | 03.64 | 01.77 |
| *TiK* | 26.27 | 25.60 |
| *Matrix* | Correction | ZAF |

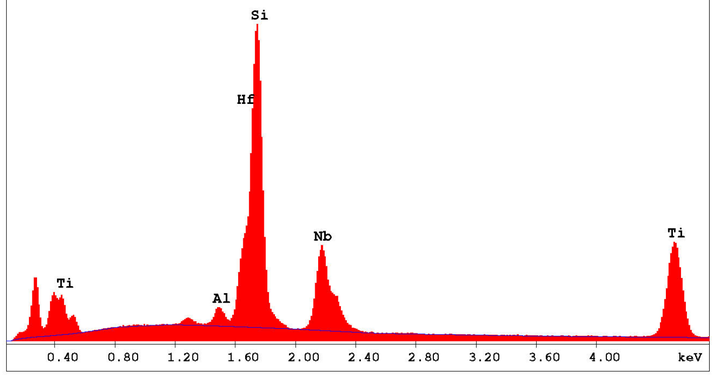



**5-4 silicide grain dark grey at edge**

**Label:**

**kV:10.0 Tilt:0.0 Take-off:34.8 Det:Octane Elect Plus**

**Res:126 Amp.T:3.84 FS:49651 Lsec:52 16-Jan-2024 11:04:52**




**EDAX ZAF Quantification (Standardless)**
**Element Normalized**
**SEC Table : Default**

| Element | Wt % | At % | K-Ratio | Z | A | F |
|---|---|---|---|---|---|---|
| AlK | 1.15 | 2.23 | 0.0111 | 1.1519 | 0.8328 | 1.0059 |
| HfM | 21.88 | 6.41 | 0.1724 | 0.7772 | 1.0134 | 1.0007 |
| SiK | 21.47 | 40.00 | 0.1988 | 1.1737 | 0.7862 | 1.0031 |
| NbL | 17.52 | 9.87 | 0.1390 | 0.9143 | 0.8662 | 1.0016 |
| TiK | 37.98 | 41.48 | 0.3809 | 1.0375 | 0.9667 | 1.0000 |
| Total | 100.00 | 100.00 | | | | |

| Element | Net Inte. | Bkgd Inte. | Inte. Error | P/B |
|---|---|---|---|---|
| AlK | 257.03 | 255.88 | 1.49 | 1.00 |
| HfM | 1182.05 | 240.51 | 0.48 | 4.91 |
| SiK | 4079.65 | 259.61 | 0.23 | 15.71 |
| NbL | 1261.52 | 203.76 | 0.45 | 6.19 |
| TiK | 1900.40 | 95.85 | 0.33 | 19.83 |

# My favorite: LLMs can help compile SEM EDS data!

Collecting spot scans creates a lot of manual transcription work from a pdf to a csv file. But we can get a LLM to read this pdf and output a csv!

*I'd like you to extract some data from a pdf file. Each page of the pdf document represents a new measurement from one of multiple samples.(There can be multiple measurements for each sample.) The sample ID, for example "5-4 silicide top" is provided at the top of each page. Each page has a table, containing the columns "Element" and "At%". Please entirely ignore the column "Wt%".*

*Please output a table in csv format. The rows should be the elements and I'd like one column for each sample. The entry in each cell should be the concentration in At%. This is science, so there is little room for creativity. It is important the the numbers are extracted exactly as they appear.*

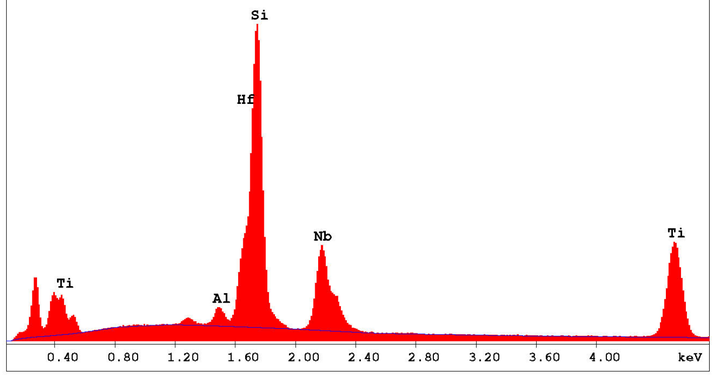
*If there is any confusion let me know! I can rephrase if the instructions aren't clear :) Thank you!*

**5-4 silicide grain dark grey at edge**

**Label:**

**kV:10.0 Tilt:0.0 Take-off:34.8 Det:Octane Elect Plus**

**Res:126 Amp.T:3.84 FS:49651 Lsec:52 16-Jan-2024 11:04:52**




**EDAX ZAF Quantification (Standardless)**
**Element Normalized**
**SEC Table : Default**

| Element | Wt % | At % | K-Ratio | Z | A | F |
|---|---|---|---|---|---|---|
| AlK | 1.15 | 2.23 | 0.0111 | 1.1519 | 0.8328 | 1.0059 |
| HfM | 21.88 | 6.41 | 0.1724 | 0.7772 | 1.0134 | 1.0007 |
| SiK | 21.47 | 40.00 | 0.1988 | 1.1737 | 0.7862 | 1.0031 |
| NbL | 17.52 | 9.87 | 0.1390 | 0.9143 | 0.8662 | 1.0016 |
| TiK | 37.98 | 41.48 | 0.3809 | 1.0375 | 0.9667 | 1.0000 |
| Total | 100.00 | 100.00 | | | | |

| Element | Net Inte. | Bkgd Inte. | Inte. Error | P/B |
|---|---|---|---|---|
| AlK | 257.03 | 255.88 | 1.49 | 1.00 |
| HfM | 1182.05 | 240.51 | 0.48 | 4.91 |
| SiK | 4079.65 | 259.61 | 0.23 | 15.71 |
| NbL | 1261.52 | 203.76 | 0.45 | 6.19 |
| TiK | 1900.40 | 95.85 | 0.33 | 19.83 |

# My favorite: LLMs can help compile SEM EDS data!

Collecting spot scans creates a lot of manual transcription work from a pdf to a csv file. But we can get a LLM to read this pdf and output a csv!

*I'd like you to extract some data from a pdf file. Each page of the pdf document represents a new measurement from one of multiple samples.(There can be multiple measurements for each sample.) The sample ID, for example "5-4 silicide top" is provided at the top of each page. Each page has a table, containing the columns "Element" and "At%". Please entirely ignore the column "Wt%".*

*Please output a table in csv format. The rows should be the elements and I'd like one column for each sample. The entry in each cell should be the concentration in At%. This is science, so there is little room for creativity. It is important the the numbers are extracted exactly as they appear.*

*If there is any confusion let me know! I can rephrase if the instructions aren't clear :) Thank you!*

| Element | 5-4 silicide top | C103 | Trialuminide 1 |
|---|---|---|---|
| **AlK** | 0.77 | 0.00 | 76.01 |
| **HfM** | 15.07 | 9.58 | 1.45 |
| **SiK** | 46.85 | 0.00 | 0.73 |
| **NbL** | 5.81 | 87.89 | 7.12 |
| **TiK** | 31.50 | 2.53 | 14.68 |

Claude 2.1 makes fewer errors than I do! (Recently, none.)

If you are worried about errors, I've noticed a tell-tale sign will be the compositions don't sum to 100%. So you can insert a sum row in the exported CSV file.



*Calculated with Claude 2.1 (free version)*

# Models allow for “system prompts” to be passed in addition to the user prompt

This feature is generally available for paid tools (e.g., $20/mo web app or API calls).

Great for automating tasks! This lets you avoid having to enter instructions every time.

This is what ChatGPT, for example, is doing in the background. Attempts to extract their system prompt have given something that is ~1,700 tokens long!

{Instructions}

+

{Text to act on}

…

{Output}

# How to maximize performance

# Good prompt design is key for good performance

Some good advice can be found on API documentation pages, for example Claudes

Some big picture tips:

1. Let the model say “I don’t know”
2. Show the model what you want (give an example)!
3. Give the model space to think
   - Ask the model to think out loud
   - Break complex tasks into simpler ones (within the same prompt, assign multiple steps to follow)
   - Chain together multiple prompts (prompt-chaining)

# Good prompt design is key for good performance

Some good advice can be found on API documentation pages, for example Claude's

Some big picture tips:

1. Let the model say “I don’t know”
2. Show the model what you want (give an example)!
3. Give the model space to think
   - Ask the model to think out loud
   - Break complex tasks into simpler ones (within the same prompt, assign multiple steps to follow)
   - Chain together multiple prompts (prompt-chaining)

*Recall…*

The best thing about AI is its ability to…

| | |
|---|---|
| learn | 4.5% |
| predict | 3.5% |
| make | 3.2% |
| understand | 3.1% |
| do | 2.9% |

Example from “What is ChatGPT Doing…and Why Does it Work? By Stephen Wolfram

The fact that, fundamentally, these models are “just” predicting the next most-likely word introduces opportunity for misbehavior

# Good prompt design is key for good performance

Some good advice can be found on API documentation pages, for example Claude's

Some big picture tips:

1. Let the model say “I don’t know”
2. Show the model what you want (give an example)!
3. Give the model space to think
   - Ask the model to think out loud
   - Break complex tasks into simpler ones (within the same prompt, assign multiple steps to follow)
   - Chain together multiple prompts (prompt-chaining)

Prompt chaining or letting the model think builds a foundation for subsequent generation. It effectively allows the model space to move toward the best answer.

*Examples:*

- Instead of asking the model to generate an essay for you, first ask it to create an outline of points to cover
- Ask the model to identify troublesome sentences in text, *then* replace them.

# Challenges using LLMs

# LLMs can hallucinate (make stuff up or get facts wrong)

*Why are refractory multi principal element alloys interesting, important, and impactful?*

Here are a few reasons why refractory multi-principal element alloys (RMEAs) are interesting, important, and impactful:

- Novel properties: RMEAs contain multiple principal elements (usually 5 or more), allowing for the development of alloys with novel properties that cannot be achieved with conventional alloys based on 1 or 2 principal elements. The high configurational entropy in RMEAs leads to unique characteristics.
- High temperature stability: RMEAs exhibit excellent high temperature strength, creep resistance, and microstructural stability…
- Oxidation and corrosion resistance: The formation of stable oxide scales imparts excellent oxidation and corrosion resistance to RMEAs at high temperatures. This expands their potential applications.



*Calculated with Claude 2.1 (free version)*

# LLMs can hallucinate (make stuff up or get facts wrong)

*Can you provide references for the role of activity on oxidation behavior?*

Certainly! Here are a couple of examples of papers published in Acta Materialia that discuss the role of activity in predicting oxidation behavior of alloys:

…

4. Xie, J., Yuan, F., Huang, W., Fan, X. & Zhuang, L. "Oxidation behavior of a high-entropy alloy CoCrNi with different carbon contents at 900 °C." Acta Materialia 140 (2017): 196-205. doi: 10.1016/j.actamat.2017.08.046.

*Truncated…*

I can’t find this paper title anywhere.

This is a real reference, but for a paper on materials selection rules for binary alloys.

And this is a real DOI but for a different paper on DFT calculations of Al-Cu!



*Calculated with GPT-3.5*

# (And sometimes, LLMs shock me with their accuracy!)

*Prompt chained request to outline then write an introduction for a paper about the oxidation of refractory MPEAs, with references…*

References:


[1] Yeh JW, et al. Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes. Adv Eng Mater, 2004. 6(5): p. 299-303.

[2] Senkov ON, et al. Microstructure and room temperature properties of a high-entropy TaNbHfZrTi alloy. J Alloys Compd, 2011. 509(41): p. 6043-6048.

[3] Gludovatz B, et al. A fracture-resistant high-entropy alloy for cryogenic applications. Science, 2014. 345(6201): p. 1153-1158.

[4] Miracle DB, Senkov ON. A critical review of high entropy alloys and related concepts. Acta Materialia, 2017. 122: p. 448-511.

[5] Gorr B, et al. High temperature oxidation behavior of refractory high entropy alloys: Effect of alloy composition. Corros Sci, 2018. 141: p. 210-219.

[6] Chen H, et al. Microstructure and oxidation behavior of new refractory high entropy alloys. J Mater Res, 2014. 29(17): p. 1903-1910.


*Calculated with Claude 2.1 (free version)*

# Models have their “ecosystem” with strengths and weaknesses

LLMs cannot directly read a pdf (yet). Parsing their data requires converting the pdf into some form of plain or rich text.


Lots of workflows for this process have been developed. Each model should have one built into its user interface, but every conversion tool behaves differently.

Each model can also remember different amounts of text, called the “context window”

**OpenAI:**

- GPT-4 Turbo: 128,000 tokens
- GPT-4: 8,192 tokens
- GPT-4 32k: 32,768 tokens

**Anthropic:**

- Claude 2: 100,000 tokens
- Claude 2.1: 200,000 tokens

**Open source models:**

- Many at ~16,000 tokens, occasionally ~32,000

# Cost and availability

**Enterprise level**

Good models aren't cheap. For example, a typical research paper is of order 10,000 tokens. This means the cost to read the paper is:

- GPT-4 turbo: $0.10
- GPT-4 32K: $0.60

Then you pay 2–3x the rate for *output* tokens. This all can be fine when your prompts are good, but the cost can add up in trial-and-error.

**Consumer level**

Plans for every provider cost $20/mo. Many graduate students could swing one of these, but paying for ≥3 becomes intractable. And being able to obtain and try multiple models *is* helpful.

**Locally**

Running models locally is a wonderful option for privacy and without worrying about experimentation. But high end computers are needed, so upfront cost is high (thousands of dollars).

**A call to PIs:**

In the context of scientific costs, these models are *effectively free*. For the cost of 1 hour of SEM time a month, your group could share access to three state-of-the-art models. These are useful *tools* and will be a bigger part of our lives moving forward.

# Looking forward…

# LLMs are getting more capable and more connected monthly

A LLM by itself is almost a gimmick. A LLM as part of a connected whole can do wonderful things.

The limitations of LLMs will fade away with time. Context windows will increase. Hallucinations will ebb.

**What's on the horizon?**

1. Retrieval augmented generation
2. Capability to describe images
3. Incorporation of LLMs into writing, integrated development environments, and note-taking tools