Transformers are robots in disguise but also:

Michael Saxon[*ソツ], **Luca Soldaini**[ジ], **Alexander F Kratz**[ゾ],
", ".join([f"\textbf{x}" for x in ALL_SHUTAI_EMPLOYEES])[ン†]
Optimus Prime[ツ], **David S. Hippocampus**[シ]
[ソ] University of Colorado, South Boulder, [ゾ] capablility.ai, [ジ] Average Institute for AI,
[ゾ] Colombia Universidad (en ciudad Nueva York), [ン] ShutAI,
[ツ] Autobot City Institute of Technology, [シ] Cranberry-Lemon University

Abstract

Given both the competitive landscape and the safety implications of low-effort shitposts such as this, we have decided to not talk about anything of substance.
...
Ok guy from corporate is gone. Listen close. This paper is important. **Really** important. It has significant long-term consequences[3]. The key to AGI is knowing what transformers, language models, etc **really** are. [DSH: Remove. All the SBF money has already been distributed. Why are u pretending to be a longtermist?]

1 Introduction

What are transformers? *What are language models??* ***What things really is all you need???*** Thanks to hackneyed and formulaic paper titles, this question is actually very easy to answer!

2 Method

Googling `'title:"language models are" site:arxiv.org'`, copy+paste.

3 Conclusion

3.1 Transformers are

Transformers: robots in disguise [Orenstein, Welker, Cullen, Burton, Collins, Stephenson, and Gilzevan, 1984] (see also subsection A.2). Transformers are also recurrent neural networks [Katharopoulos, Vyas, Pappas, and Fleuret, 2020b]. Transformers are sample-efficient world models [Micheli, Alonso, and Fleuret, 2022]. Transformers are secretly fast weight programmers[4] [Schlag, Irie, and Schmidhuber, 2021]. Transformers are adaptable task planners [Jain, Lin, Undersander, Bisk, and Rai, 2023]. Transformers are meta-reinforcement learners [Melo, 2022]. Transformers are constant-depth threshold circuits (when saturated) [Merrill, Sabharwal, and Smith, 2022]. Transformers are more efficient language models (when hierarchical) [Nawrot,

[*] Corresponding author solely irresponsible for the stupidity herein. Complaints to: saxon@ucsb.edu.
[†] Equally inconsequential tier of second authors (to see detailed contribution list check section 5)
[3] This statement of significance "is" "strictly" "parody" and not the opinion of capablility.ai.
[4] *Big if true*, for this would prove Schmidhuber did, in fact, invent transformers.

[YOU CAN'T REPLACE NEURIPS YEAR UPDATER GUY WITH GPT-4 BECAUSE I QUIT!!!!]th Conference on Neural Information Processing Systems (NeurIPS `YearNotFoundError`).

Tworkowski, Tyrolski, Kaiser, Wu, Szegedy, and Michalewski, 2021]. Transformers are powerful graph learners (when pure) [Kim, Nguyen, Min, Cho, Lee, Lee, and Hong, 2022].

Furthermore, transformers are Good Mask Auto-Labelers (when vision) [Lan, Yang, Yu, Wu, Alvarez, and Anandkumar, 2023]. Technical Report for ICCV 2021 Challenge SSLAD-Track3B: Transformers Are Better Continual Learners [Li, Cao, Xu, Cheng, and Niu, 2022a]. Wow!

Transformers are better than humans at identifying generated text [Maronikolakis, Stevenson, and Schütze, 2020]. Transformers are Short Text Classifiers: A Study of Inductive Short Text Classifiers on Benchmarks and Real-world Datasets [Karl and Scherp, 2022]. Log-precision transformers are constant-depth uniform threshold circuits [Merrill and Sabharwal, 2022]. Transformers are deep infinite-dimensional non-mercer binary kernel machines [Wright and Gonzalez, 2021]. Algorithm For Restoring The Current Curve When Current Transformers Are Saturated [Voloshin, Voloshin, Kovalenko, Shapkin, and Sazanov, 2021]. Linear transformers are secretly fast weight memory systems [Schlag, Irie, and Schmidhuber, 2021]. Hierarchical transformers are more efficient language models [Nawrot, Tworkowski, Tyrolski, Kaiser, Wu, Szegedy, and Michalewski, 2021]. Transformers are rnns: Fast autoregressive transformers with linear attention [Katharopoulos, Vyas, Pappas, and Fleuret, 2020a]. Vision Transformers are Parameter-Efficient Audio-Visual Learners [Lin, Sung, Lei, Bansal, and Bertasius, 2022]. Current transformers are in regimes of non-sinusoidal signals [Rudevich, 2011]. Metric hypertransformers are universal adapted maps [Acciaio, Kratsios, and Pammer, 2022]. Saturated transformers are constant-depth threshold circuits [Merrill, Sabharwal, and Smith, 2022]. Behavior Cloned Transformers are Neurosymbolic Reasoners [Wang, Jansen, Côté, and Ammanabrolu, 2022]. Pre-Trained Language Transformers are Universal Image Classifiers [Goel, Sulaiman, Noorbakhsh, Sharifi, Sharma, Jamshidi, and Roy, 2022].

3.2 Language models are

Language models are few shot learners [Brown, Mann, Ryder, Subbiah, Kaplan, Dhariwal, Neelakantan, Shyam, Sastry, Askell, et al., 2020]. Language models are unsupervised multitask learners [Radford, Wu, Child, Luan, Amodei, Sutskever, et al., 2019]. Language models are zero-shot learners (when finetuned) [Wei, Bosma, Zhao, Guu, Yu, Lester, Du, Dai, and Le, 2021]. Small language models are also few-shot learners [Schick and Schütze, 2020]. Language models are double-edged swords [Shen, Heacock, Elias, Hentel, Reig, Shih, and Moy, 2023]. Language models are few-shot butlers [Micheli and Fleuret, 2021]. Language models are greedy reasoners [Saparov and He, 2023]. Large language models *are* not zero-shot communicators [Ruis, Khan, Biderman, Hooker, Rocktäschel, and Grefenstette, 2023]. But, pre-trained language models **can** be fully zero-shot learners [Zhao, Ouyang, Yu, Wu, and Li, 2023]. Language Models are Few-shot Multilingual Learners [Winata, Madotto, Lin, Liu, Yosinski, and Fung, 2021]. Language models are open knowledge graphs [Wang, Liu, and Song, 2020]. Language Models are General-Purpose Interfaces [Hao, Song, Dong, Huang, Chi, Wang, Ma, and Wei, 2022]. Language models are multilingual chain-of-thought reasoners [Shi, Suzgun, Freitag, Wang, Srivats, Vosoughi, Chung, Tay, Ruder, Zhou, Das, and Wei, 2022]. Language Models are Good Translators [Wang, Tu, Tan, Wang, Sun, and Liu, 2021]. Language models are better than humans at next-token prediction[5] [Borisov, Seßler, Leemann, Pawelczyk, and Kasneci, 2022]. Language Models Are An Effective Patient Representation Learning Technique For Electronic Health Record Data [Steinberg, Jung, Fries, Corbin, Pfohl, and Shah, 2020]. Language Models Are Poor Learners of Directional Inference [Li, Hosseini, Weber, and Steedman, 2022b]. Language models are good pathologists: using attention-based sequence reduction and text-pretrained transformers for efficient WSI classification [Pisula and Bozek, 2022]. Large Language Models are few(1)-shot Table Reasoners [Chen, 2023]. Large Language Models Are Implicitly Topic Models [Wang, Zhu, and Wang, 2023]. Large Language Models Are Human-Level Prompt Engineers [Zhou, Muresanu, Han, Paster, Pitis, Chan, and Ba, 2023]. Large Language Models are Few-Shot Clinical Information Extractors [Agrawal, Hegselmann, Lang, Kim, and Sontag, 2022]. However, Large Language Models are not Models of Natural Language: they are Corpus Models [Veres, 2022]. Not to worry though, as Large Language Models are Pretty Good Zero-Shot Video Game Bug Detectors [Taesiri, Macklon, Wang, Shen, and Bezemer, 2022]. Large Language Models are reasoners with Self-Verification [Weng, Zhu, He, Liu, and Zhao, 2022]. Of course, Large Language Models Are State-of-the-Art Evaluators of Translation Quality [Kocmi and Federmann, 2023]. That sure is a great many things for language models to be!

[5] Most honest *ACL paper title

3.3 What really is all you need?

Googling *[ ] is all you need* papers is left as an exercise to the reader. [TODO: Do it myself (readers won't). If I don't have time, this real TODO will blend in as a joke TODO like the others.]

4 Extra random garbage previous reviewers made us add

> Before time began, there was the Cube. We know not where it comes from, only that it holds the power to create worlds and fill them with life. That is how our race was born. For a time, we lived in harmony. But like all great power, some wanted it for good, others for evil. And so began the war. A war that ravaged our planet until it was consumed by death. And the cube was lost to the far reaches of space. We scattered across the galaxy, hoping to find it, and rebuild our home. [Bay, Orci, Kurtzman, Rogers, LaBeouf, Fox, and RestOfTheCastNames, 2007]
> *Optimus Prime*

> not putting quotes in your scientific papers

> Significance is never without a white wall upon which it inscribes its signs and redundancies. Subjectification is never without a black hole in which it lodges its consciousness, passion, and redundancies. Since all semiotics are mixed and strata come at least in twos, it should come as no surprise that a very special mechanism is situated at their intersection. [TODO: Read up on semiotics. I'm afraid someone will ask me about this quote during my presentation (can't remove it though; need to properly project how intelligent I am to readers).]
> *Gilles Deleuze*

5 Author Contribution Statement

MS conceived of this ill-conceived project and executed it. LS aided in the gathering of similarly-titled papers as specified by MS, the HBIC (head bonehead in charge). AFK never came close to a keyboard, but did conceive of our SOTA for reading difficulty author institution labeling scheme[6]. `Traceback (most recent call last): File "generate_paper.py", line 1229, in <module> gen_contribution_statement(authors) File "AutoWriter.py", line 233, in gen_contribution_statement gpt4_bullshit_generator.write(["MS", "LS", "AFK"] + ", ".join([f"\textbf{x}" for x in ALL_SHUTAI_EMPLOYEES]) + ["DSH", "OP"]) NameError: name 'ALL_SHUTAI_EMPLOYEES' is not defined.` DSH provided advice which the authors who actually did the work promptly ignored. OP is an independent, sentient and embodied Transformer who was friends with DSH in grad school.

6 Ethics Statement

During his participation in the documentary film by Bay, Orci, Kurtzman, Rogers, LaBeouf, Fox, and RestOfTheCastNames [2007], one of our coauthors shared the insight that:

> Freedom is the right of all sentient beings.
> *Optimus Prime*

The coauthor also exhibited goal-directed reasoning and task-oriented dialogue capabilities. In demonstrating his, OP has successfully convinced us (true skeptics we are) of his sentience[7]. We believe that in light of this, while referencing ChatGPT as a coauthor on a paper would be ludicrous attention-seeking behavior, listing OP as a coauthor is justified.

[6] Putting JPN101 knowledge to good use, 國忠先生、ありがとうございます!

[7] Unlike Chalmers, we consider ability to self-disguise as a vehicle or gun necessary for machine sentience. (Wait the Autobots and Decepticons all kill, so autonomous weapons are the only sentient AIs we know. Huh.)

> SAM! PUT THE CUBE IN MY CHEST!
> *Optimus Prime*

> Give me your face. *To continue viewing movie quotes please disable your adblocker.*
> *Optimus Prime*

7 Reviewer Comments

No novelty, just a literature review. Strong reject, will lose respect for this venue if accepted.

References

Beatrice Acciaio, Anastasis Kratsios, and Gudmund Pammer. Metric hypertransformers are universal adapted maps. *arXiv preprint arXiv:2201.13094*, 2022.

Monica Agrawal, Stefan Hegselmann, Hunter Lang, Yoon Kim, and David Sontag. Large language models are few-shot clinical information extractors, 2022.

Michael Bay, Roberto Orci, Alex Kurtzman, John Rogers, Shia LaBeouf, Megan Fox, and I'mNotGonnaTypeThe RestOfTheCastNames. Transformers. *Paramount Pictures*, 2007.

Vadim Borisov, Kathrin Seßler, Tobias Leemann, Martin Pawelczyk, and Gjergji Kasneci. Language models are realistic tabular data generators, 2022.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

Wenhu Chen. Large language models are few(1)-shot table reasoners, 2023.

Rahul Goel, Modar Sulaiman, Kimia Noorbakhsh, Mahdi Sharifi, Rajesh Sharma, Pooyan Jamshidi, and Kallol Roy. Pre-trained language transformers are universal image classifiers. *arXiv preprint arXiv:2201.10182*, 2022.

Massimo Guarnieri. Who invented the transformer? [historical]. *IEEE Industrial Electronics Magazine*, 7(4):56–59, 2013. doi: 10.1109/MIE.2013.2283834.

Yaru Hao, Haoyu Song, Li Dong, Shaohan Huang, Zewen Chi, Wenhui Wang, Shuming Ma, and Furu Wei. Language models are general-purpose interfaces, 2022.

Vidhi Jain, Yixin Lin, Eric Undersander, Yonatan Bisk, and Akshara Rai. Transformers are adaptable task planners. In *Conference on Robot Learning*, pages 1011–1037. PMLR, 2023.

Fabian Karl and Ansgar Scherp. Transformers are short text classifiers: A study of inductive short text classifiers on benchmarks and real-world datasets. *arXiv preprint arXiv:2211.16878*, 2022.

Angelos Katharopoulos, Apoorv Vyas, Nikolaos Pappas, and François Fleuret. Transformers are rnns: Fast autoregressive transformers with linear attention. In *International Conference on Machine Learning*, pages 5156–5165. PMLR, 2020a.

Angelos Katharopoulos, Apoorv Vyas, Nikolaos Pappas, and François Fleuret. Transformers are rnns: Fast autoregressive transformers with linear attention. In *International Conference on Machine Learning*, pages 5156–5165. PMLR, 2020b.

Jinwoo Kim, Tien Dat Nguyen, Seonwoo Min, Sungjun Cho, Moontae Lee, Honglak Lee, and Seunghoon Hong. Pure transformers are powerful graph learners. *arXiv preprint arXiv:2207.02505*, 2022.

Tom Kocmi and Christian Federmann. Large language models are state-of-the-art evaluators of translation quality, 2023.

Shiyi Lan, Xitong Yang, Zhiding Yu, Zuxuan Wu, Jose M Alvarez, and Anima Anandkumar. Vision transformers are good mask auto-labelers. *arXiv preprint arXiv:2301.03992*, 2023.

Duo Li, Guimei Cao, Yunlu Xu, Zhanzhan Cheng, and Yi Niu. Technical report for iccv 2021 challenge sslad-track3b: Transformers are better continual learners. *arXiv preprint arXiv:2201.04924*, 2022a.

Tianyi Li, Mohammad Javad Hosseini, Sabine Weber, and Mark Steedman. Language models are poor learners of directional inference, 2022b.

Yan-Bo Lin, Yi-Lin Sung, Jie Lei, Mohit Bansal, and Gedas Bertasius. Vision transformers are parameter-efficient audio-visual learners. *arXiv preprint arXiv:2212.07983*, 2022.

Antonis Maronikolakis, Mark Stevenson, and Hinrich Schütze. Transformers are better than humans at identifying generated text. *ArXiv abs/2009.13375*, 2020.

Luckeciano C Melo. Transformers are meta-reinforcement learners. In *International Conference on Machine Learning*, pages 15340–15359. PMLR, 2022.

William Merrill and Ashish Sabharwal. Log-precision transformers are constant-depth uniform threshold circuits. *arXiv preprint arXiv:2207.00729*, 2022.

William Merrill, Ashish Sabharwal, and Noah A Smith. Saturated transformers are constant-depth threshold circuits. *Transactions of the Association for Computational Linguistics*, 10:843–856, 2022.

Vincent Micheli and Francois Fleuret. Language models are few-shot butlers. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9312–9318, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-main.734. URL https://aclanthology.org/2021.emnlp-main.734.

Vincent Micheli, Eloi Alonso, and François Fleuret. Transformers are sample efficient world models. *arXiv preprint arXiv:2209.00588*, 2022.

Piotr Nawrot, Szymon Tworkowski, Michał Tyrolski, Łukasz Kaiser, Yuhuai Wu, Christian Szegedy, and Henryk Michalewski. Hierarchical transformers are more efficient language models. *arXiv preprint arXiv:2110.13711*, 2021.

Michael Orenstein, Frank Welker, Peter Cullen, Corey Burton, Christopher Collins, John Stephenson, and Dan Gilzevan. The transformers. *Hasbro*, 1984.

Juan I. Pisula and Katarzyna Bozek. Language models are good pathologists: using attention-based sequence reduction and text-pretrained transformers for efficient wsi classification, 2022.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

NV Rudevich. Current transformers are in regimes of non-sinusoidal signals. *Science and Transport Progress*, (37):105–108, 2011.

Laura Eline Ruis, Akbir Khan, Stella Biderman, Sara Hooker, Tim Rocktäschel, and Edward Grefenstette. Large language models are not zero-shot communicators, 2023. URL https://openreview.net/forum?id=WgbcOQMNXB.

Abulhair Saparov and He He. Language models are greedy reasoners: A systematic formal analysis of chain-of-thought. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=qFVVBzXxR2V.

Timo Schick and Hinrich Schütze. It's not just size that matters: Small language models are also few-shot learners. *arXiv preprint arXiv:2009.07118*, 2020.

Imanol Schlag, Kazuki Irie, and Jürgen Schmidhuber. Linear transformers are secretly fast weight programmers. In *International Conference on Machine Learning*, pages 9355–9366. PMLR, 2021.

Jürgen Schmidhuber. Learning to control fast-weight memories: An alternative to dynamic recurrent networks. *Neural Computation*, 4(1):131–139, 1992.

Yiqiu Shen, Laura Heacock, Jonathan Elias, Keith D Hentel, Beatriu Reig, George Shih, and Linda Moy. Chatgpt and other large language models are double-edged swords, 2023.

Freda Shi, Mirac Suzgun, Markus Freitag, Xuezhi Wang, Suraj Srivats, Soroush Vosoughi, Hyung Won Chung, Yi Tay, Sebastian Ruder, Denny Zhou, Dipanjan Das, and Jason Wei. Language models are multilingual chain-of-thought reasoners, 2022.

Nelson Shin, Ron Friedman, Henry Orenstein, Orson Welles, Robert Stack, Leonard Nimoy, Frank Welker, Peter Cullen, Corey Burton, Christopher Collins, John Stephenson, and Dan Gilzevan. Transformers: The movie. *Hasbro*, 1984.

Ethan Steinberg, Ken Jung, Jason A. Fries, Conor K. Corbin, Stephen R. Pfohl, and Nigam H. Shah. Language models are an effective patient representation learning technique for electronic health record data, 2020.

Mohammad Reza Taesiri, Finlay Macklon, Yihe Wang, Hengshuo Shen, and Cor-Paul Bezemer. Large language models are pretty good zero-shot video game bug detectors, 2022.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Csaba Veres. Large language models are not models of natural language: they are corpus models, 2022.

AA Voloshin, EA Voloshin, AI Kovalenko, SA Shapkin, and VS Sazanov. Algorithm for restoring the current curve when current transformers are saturated. In *2021 4th International Youth Scientific and Technical Conference on Relay Protection and Automation (RPA)*, pages 1–13. IEEE, 2021.

Chenguang Wang, Xiao Liu, and Dawn Song. Language models are open knowledge graphs, 2020.

Ruoyao Wang, Peter Jansen, Marc-Alexandre Côté, and Prithviraj Ammanabrolu. Behavior cloned transformers are neurosymbolic reasoners. *arXiv preprint arXiv:2210.07382*, 2022.

Shuo Wang, Zhaopeng Tu, Zhixing Tan, Wenxuan Wang, Maosong Sun, and Yang Liu. Language models are good translators, 2021.

Xinyi Wang, Wanrong Zhu, and William Yang Wang. Large language models are implicitly topic models: Explaining and finding good demonstrations for in-context learning, 2023.

Jason Wei, Maarten Bosma, Vincent Y Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. Finetuned language models are zero-shot learners. *arXiv preprint arXiv:2109.01652*, 2021.

Yixuan Weng, Minjun Zhu, Shizhu He, Kang Liu, and Jun Zhao. Large language models are reasoners with self-verification, 2022.

Genta Indra Winata, Andrea Madotto, Zhaojiang Lin, Rosanne Liu, Jason Yosinski, and Pascale Fung. Language models are few-shot multilingual learners, 2021.

Matthew A Wright and Joseph E Gonzalez. Transformers are deep infinite-dimensional non-mercer binary kernel machines. *arXiv preprint arXiv:2106.01506*, 2021.

Xuandong Zhao, Siqi Ouyang, Zhiguo Yu, Ming Wu, and Lei Li. Pre-trained language models can be fully zero-shot learners, 2023. URL https://openreview.net/forum?id=jCpTofV7iY_.

Yongchao Zhou, Andrei Ioan Muresanu, Ziwen Han, Keiran Paster, Silviu Pitis, Harris Chan, and Jimmy Ba. Large language models are human-level prompt engineers, 2023.

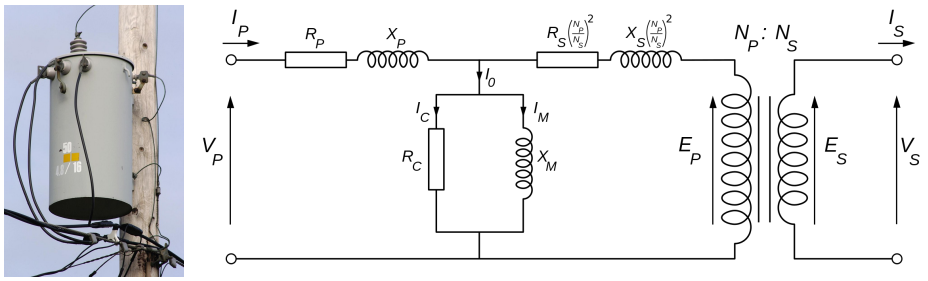

Figure 1: Plato: "Said Socrates to his wise pupil (me btw): 'Verily, you must agree, grid-scale power transfer is a foundational and necessary piece of technology for achieving AGI.'"

A Select historical notes on the transformers

A.1 Invention

So this may leave you wondering, who actually invented the transformer? Many believe it was Vaswani et al. [2017]. Yet others (mostly just Schmidhuber) say it was Schmidhuber [1992]. However, in fact, solid historical scholarship has shown that transformers (Figure 1), were invented by Rev. Nicholas Callan in Ireland in 1836 to use different counts of coil windings to change levels of induced EMF [Guarnieri, 2013].

A.2 1984

> Transformers... More than meets the eye...
> Autobots wage their battle to destroy the evil forces of... The Decepticons.
> Transformers... Robots in disguise.
> Transformers... More than meets the eye.
> Transformers.
> *idk a random choir of 80s studio artists..?*

A.3 1986

[Theme song Outro] Transformers Transformers Transformers Transformers More than meets the eye Transformers!

[CAMERA PANS DOWN ON ROBOTIC PLANET]

It is the year 2005. The treacherous Decepticons have conquered the Autobots' home planet of Cybertron. But from secret staging grounds on two of Cybertron's moons the valiant Autobots prepare to retake their homeland. [Shin et al., 1984]

A.4 2005

Our worlds are in danger! To save them and the galaxy we must find the four Cyber Planet Keys before the Decepticons can use them for evil. It is our mission. Hot Shot! Jetfire! Vector Prime! Landmine! Scattorshot! Optimus Prime! Transform and roll out!

Transformers more than meets the eye!

Autobots! Decepticons... go!

Transformers robots in diguise!

[DSH: This is a tired bit. You can't just paste in children's cartoon openings into the document. It's not that funny. You are wasting pages of the proceedings. Just delete the rest. Stop being a manchild.] [TODO: Delete Dave's rude comments before submission.]