

The paper "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" by Emily Bender and others is extensive, and summarizing each section into a slide would require a concise approach while still capturing the essence of the content. Here are brief summaries for each section:

0. Abstract:

Highlights the trend of increasingly large language models in NLP, discussing risks such as environmental, financial, and societal impacts, and proposes recommendations for mitigation.

1. Introduction:

Focuses on the rising size of language models, questioning the associated risks and strategies for mitigating these risks, including environmental and financial concerns.

2. Background:

Provides an overview of the evolution of language modeling, from n-gram models to neural models and Transformers, and discusses concerns related to data size and representativeness.

3. Environmental and Financial Cost:

Discusses the high environmental and financial costs of training large models, highlighting implications for accessibility, equity in research, and environmental impact.



#### 4. Unfathomable Training Data:

Addresses the challenges of large datasets in capturing diverse perspectives, risks of encoding biases, and reinforcing hegemonic viewpoints.



Section 4 of the paper "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" by Emily Bender and others, titled "Unfathomable Training Data," addresses several key issues:

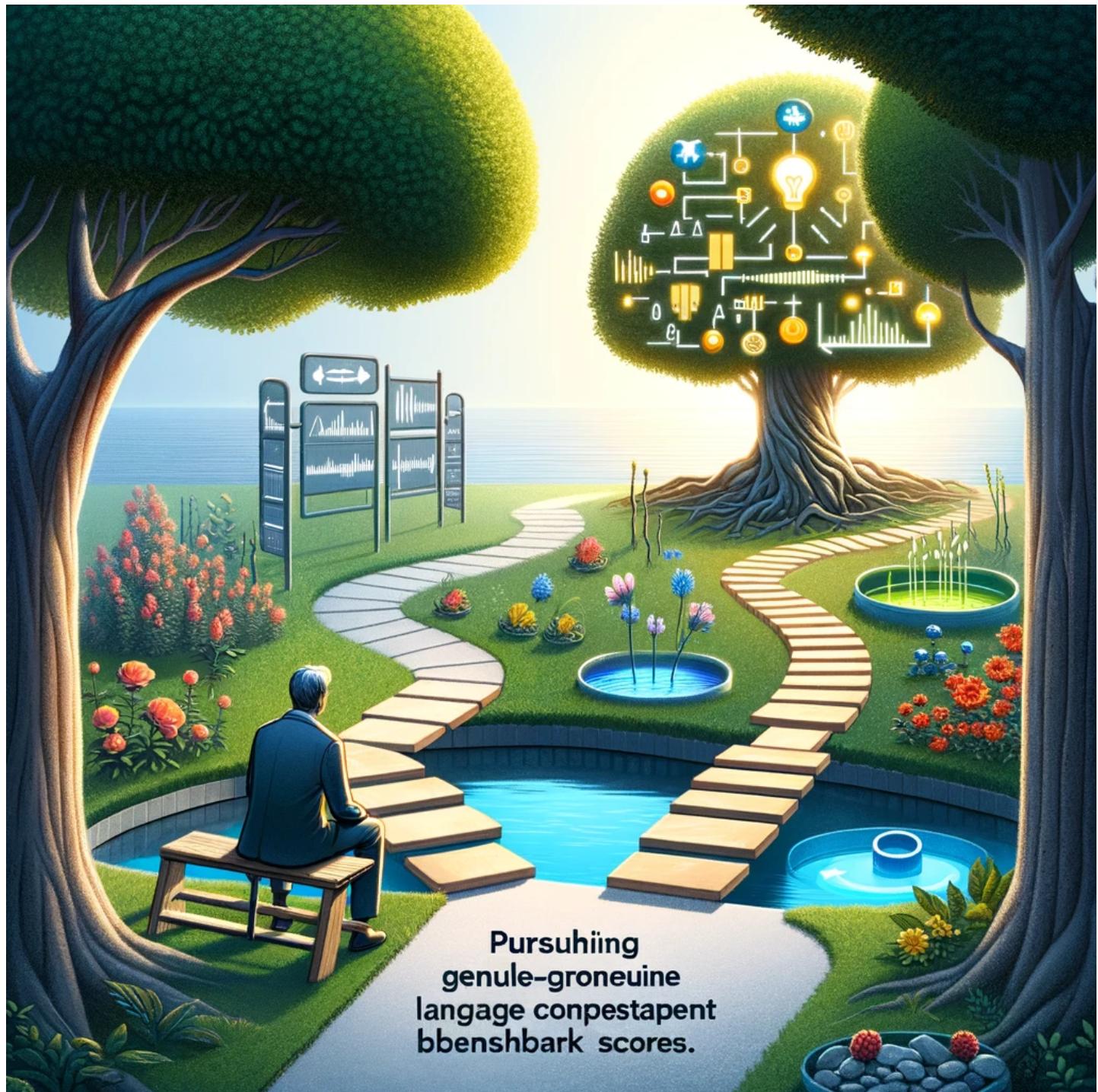
1. **Size Doesn't Guarantee Diversity:** Large datasets do not necessarily capture diverse perspectives. They often reflect dominant, hegemonic views, leading to the marginalization of other voices.
2. **Static Data/Changing Social Views:** The static nature of training data can lag behind rapidly evolving social views and languages, resulting in outdated or biased representations.
3. **Encoding Bias:** Large language models tend to encode and amplify biases present in their training data, affecting their outputs and potentially causing harm.

4. Curation, Documentation, & Accountability: The importance of thorough curation and documentation of datasets is emphasized, highlighting the need for transparency and accountability in model training.

These points underscore the complexity and potential pitfalls of relying on massive, uncurated datasets for training language models.

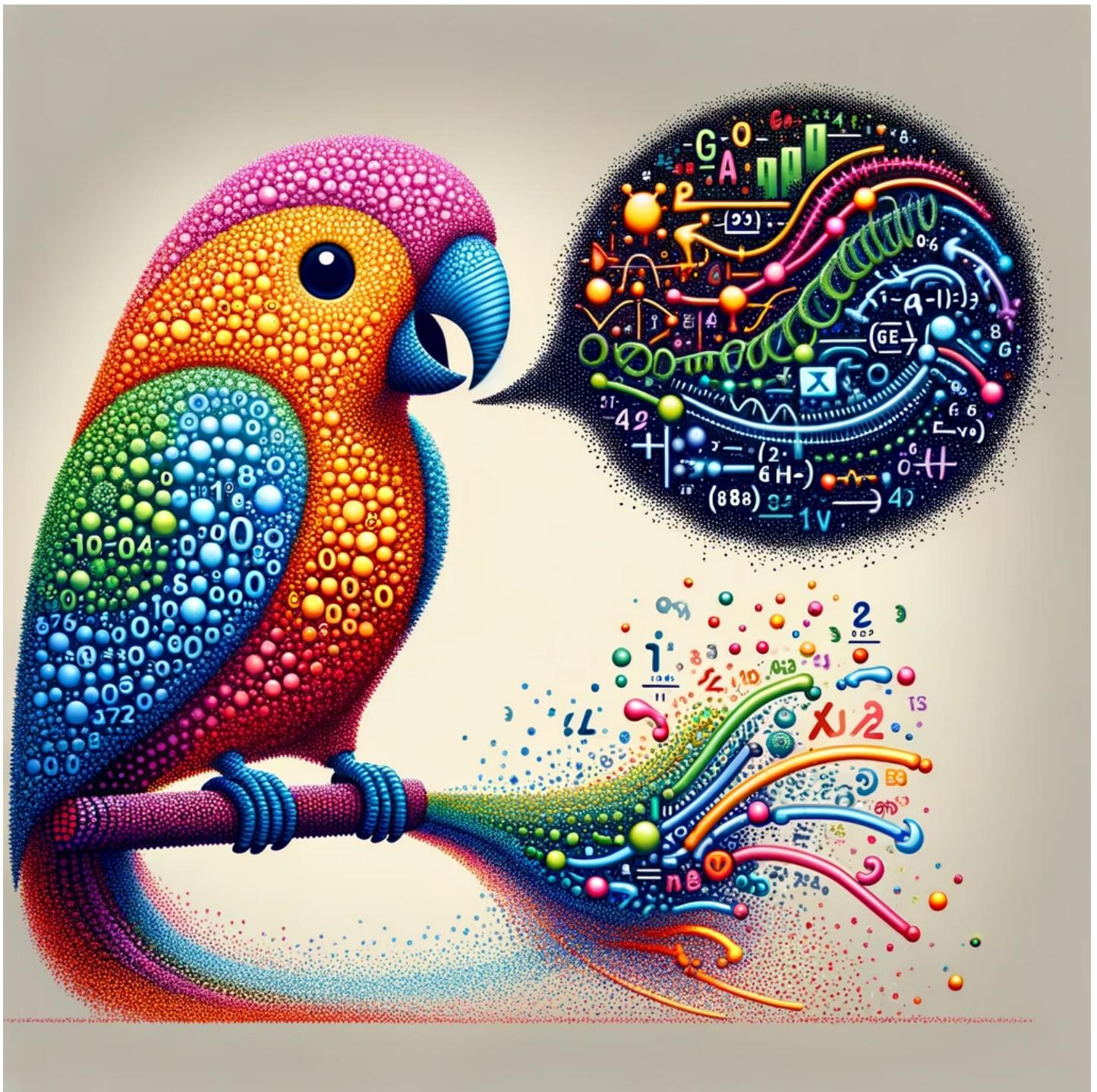
5. Down the Garden Path:

Focuses on misdirected research efforts, emphasizing concerns about overemphasis on benchmarks without true language understanding.



6. Stochastic Parrots:

Explores the tendency of language models to reproduce biases from training data, discussing potential real-world harm and misinformation.



## 7. Paths Forward:

Provides recommendations for responsible language model development, emphasizing careful planning and consideration of broader impacts.









## 8. Conclusion:

Summarizes the risks associated with large language models and calls for a balanced approach in their development, considering societal impacts.

For a more detailed summary and inclusion in your presentation, you may need to reference specific parts of the paper to extract key points and examples used in each section.



