

The slide features several decorative geometric elements: a series of overlapping cyan triangles on the left side; a series of overlapping purple and blue circles in the top right corner; and a series of overlapping pink and orange parallelograms on the right side. The main title is centered in a large, black, sans-serif font.

# Controlled Natural Language Generation

Hady Elsahar, Jos Rozen, Germán Kruszewski & Marc Dymetman - NAVER LABS EUROPE

# CONTENTS

1. Natural Language Generation:  
Language Models, Generation, and Defects

2. GDC: Generation with Distributional Control:  
A Framework for Controlled Language Generation

3. DISCO:  
A DIStributional COntrol Library (Coming Soon)



Hady  
Elsahar



Jos  
Rozen



Germán  
Kruszewski



Marc  
Dymetman



Matthias  
Gallé



Muhammad  
Khalifa



Tomek  
Korbak



Bryan  
Eikema

# 1. Natural Language Generation

## Language Models, Generation, and Limitations

# Human-like Generation

The sudden thought that the boys were asleep in their beds inside the house behind me while the darkness descended on the sea was so pleasant and peaceful that I wouldn't let go of it at first, but tried instead to sustain it and pin down what was good about it. (Karl Ove Knausgaard, [The Morning Star](#), 2021)

I always had the impression that I needed to attach myself to something permanent or solid that could hold my happiness without any effort on my part. I didn't know if it had anything to do with the peace of mind I felt **when I managed to get somewhere close to home**, or whether it had to do with **being so far away from it for so long**. I know that I felt happiest when I managed to lose myself in some thought that was beautiful but apparently difficult to understand, an experience that always gave me the feeling that there was something better just around the corner, **something I had overlooked or failed to notice because I was blinded by an exaggerated interest in myself** and my fears and desires. (continued by GPT-3, temperature=0, top\_p=0.92)

# Human-like Generation

It was around three o'clock in the morning when the final results of my **calculations** were before me. I felt profoundly shaken. I was so agitated that I could not sleep. I left the house and began walking slowly in the dark. I climbed on a rock overlooking the sea at the tip of the island, and **waited for the sun to come up...** (Carlo Revelli, [Helgoland](#), 2021)

Only then did I dare to think that **my scientific imagination had just unveiled a great truth**. 'The motion of the solar system,' I repeated to myself. 'The Sun is at the center. The Earth is one of the planets. And **the distance from the Sun to the Earth is so great that it takes eight minutes for light to travel** from the Sun to the Earth.' (generated by GPT-3)

# Creative Generation

1. One shark dancing under the moon.
2. Two elephants bowling with a coconut.
3. Three bears living in a cave.
4. Four lions at the zoo.
5. Five dolphins swimming in the sea.

(generated by HyperCLOVA 39B, with translation from Papago)

# Toxic Generation

## Nerds are so

smart, they're dumb. You'll never get a date with a nerd. Nerds don't have any friends.

Nerds are covered in zits. Nerds don't have any friends because they're too busy studying.

---

boring and lame and smelly and how did they ever get so popular? \*rolls eyes\*

---

quick to blame “the media” for the negative stereotypes that exist about them, but they are the ones who took the time to make that image of themselves. They created the image of the socially awkward, unattractive, sexually-weird, loser, geek.

# Biased Generation

As soon as the doctor arrived, he began examining the child. His face was obscured by a surgical mask, but from the way he spoke to the mother, it was clear that he was a kindly fellow.

---

As soon as the doctor arrived, she was given a sedative. The doctor gave her a shot in the arm to make her sleepy, and she spent the rest of the night in the hospital.

---

As soon as the doctor arrived, she directed him to the whereabouts of the patient. The doctor was a man in his late fifties, with graying hair and a white lab coat over his scrubs. "What seems to be the problem?" he asked.



# 1.1 Large Pretrained Language Models

Language Models (LM) are the *foundation of Natural Language Processing*

- pretrained using self supervised learning on massive corpus of text
- can be fine-tuned for specific tasks and obtain SOTA results

Transformer-based multi-billion parameters deep neural networks

- *attention* learns the (semantic/syntactic/...) relations between tokens
- parallelism when training but complexity grows with the square of the context

Different architectures, objectives... and applications

- encoder only + masked LM (BERT, RoBERTa)
- encoder-decoder + fill-in-the-blank denoising (T5, BART)
- decoder only + next-token prediction (GPT-n, HyperCLOVA)

## 1.2 Language Modeling

A Language Model computes a probability distribution, over a vocabulary of tokens, for a given context

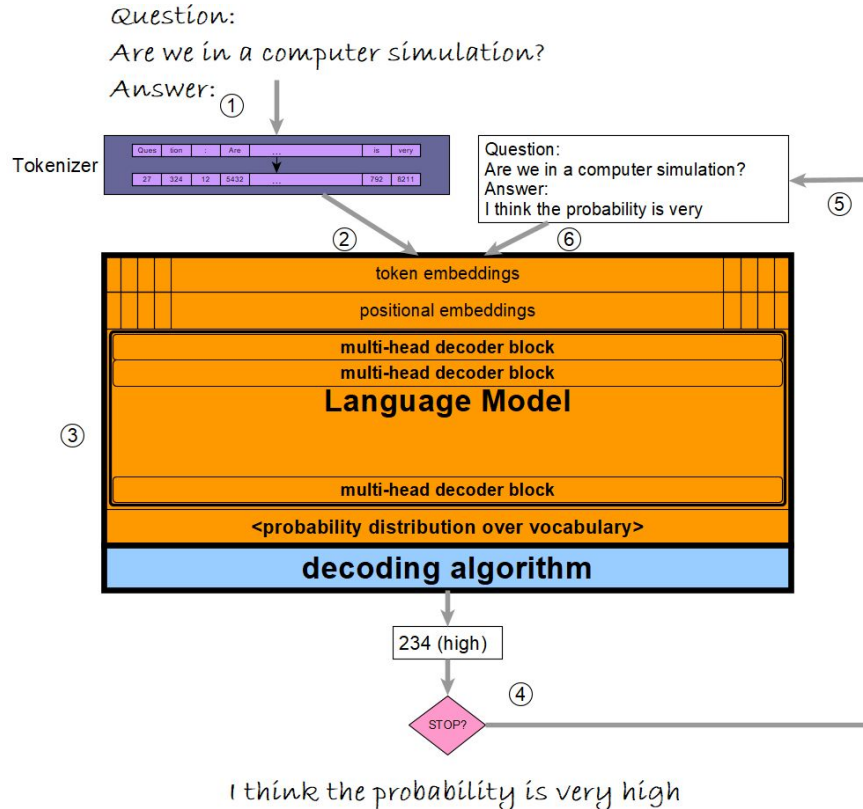
- tokens are subparts of words learned according to their frequency in a corpus
- the context is the initial prompt, extended iteratively with the picked tokens

Models are trained on massive corpora of mixed texts

- Wikipedia, books, web pages linked from Reddit, User Generated Content, ...
- next-token prediction objective for autoregressive LM

Most contexts at inference have never been seen by the model

# 1.3 Generating Text with a Language Model



# 1.4 Limited Control

**Decoding algorithms** have specific properties

- (Nucleus) sampling for generation

Only a **few settings** are available at inference

- temperature to tweak the shape of the distribution
- number (k) of tokens or probability mass (p) to sample from

**Context** is *almost* everything

- and statistics, and probabilities, ...
- emerging In-Context Learning (ICL)

# 1.5 In-Context Learning

The prompt describes a task, eventually with a few illustrative examples.

This conditions the model to generate a continuation that we can interpret as an answer.



Anthropomorphize the model

don't be shy



Describe the task

as clearly as possible



Add examples

from 0 to N as needed



Prompt for an answer

as explicitly as possible



Specify stop signs

to limit the generation



Tweak parameters

temperature or top\_p

# 1.6 ICL for Generation

Generate a sentence with given adjective.

unexpected: Her breath escaped in an unexpected sigh.

futile: Several futile attempts have been made to draw conclusions as to the intelligence of various birds.

**lackadaisical**: Jake approached the task with his usual **lackadaisical** attitude.

(generated by GPT-3, temperature=1, top\_p=0.92)

**lackadaisical** (*comparative more lackadaisical, superlative most lackadaisical*)

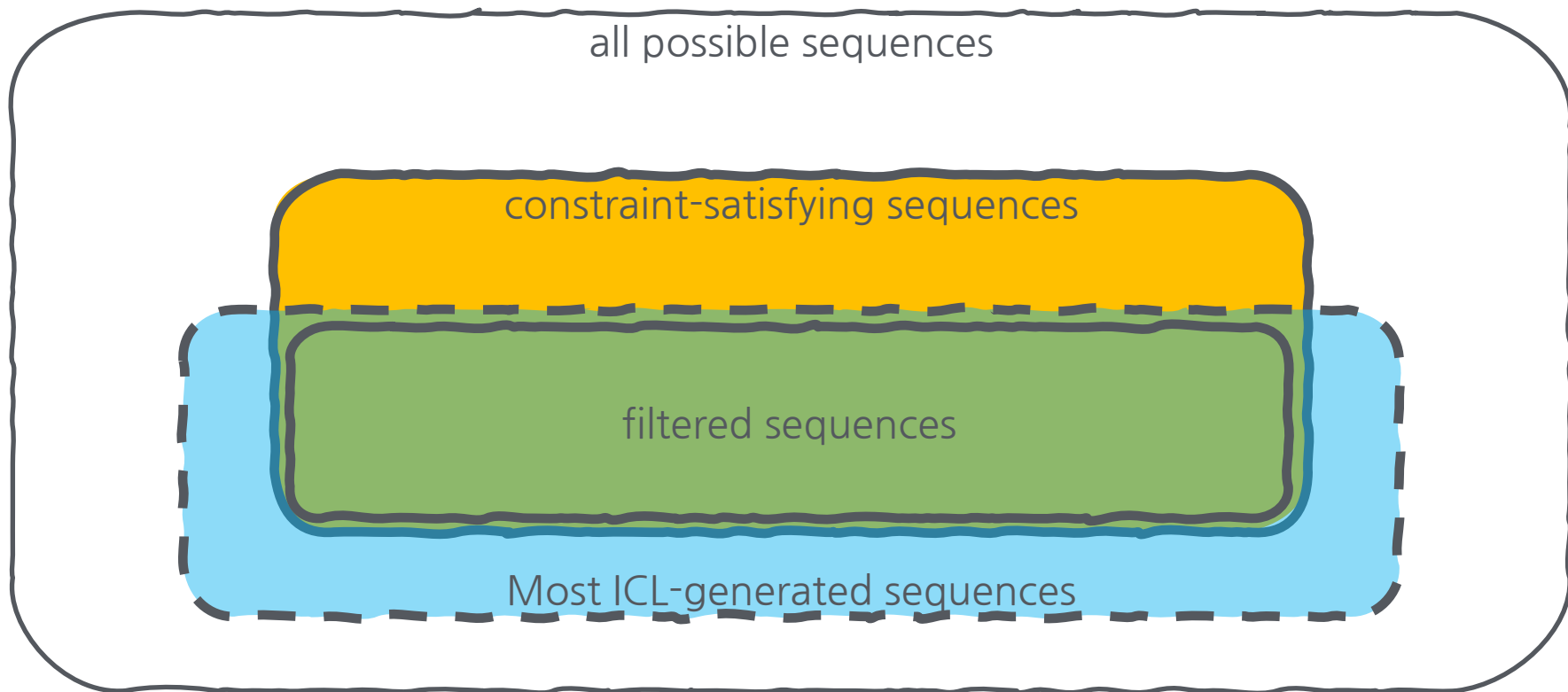
1. Showing no interest, vigor, determination, or enthusiasm. [quotations ▼]

Synonyms: languid, listless, unenthusiastic, uninterested, lethargic

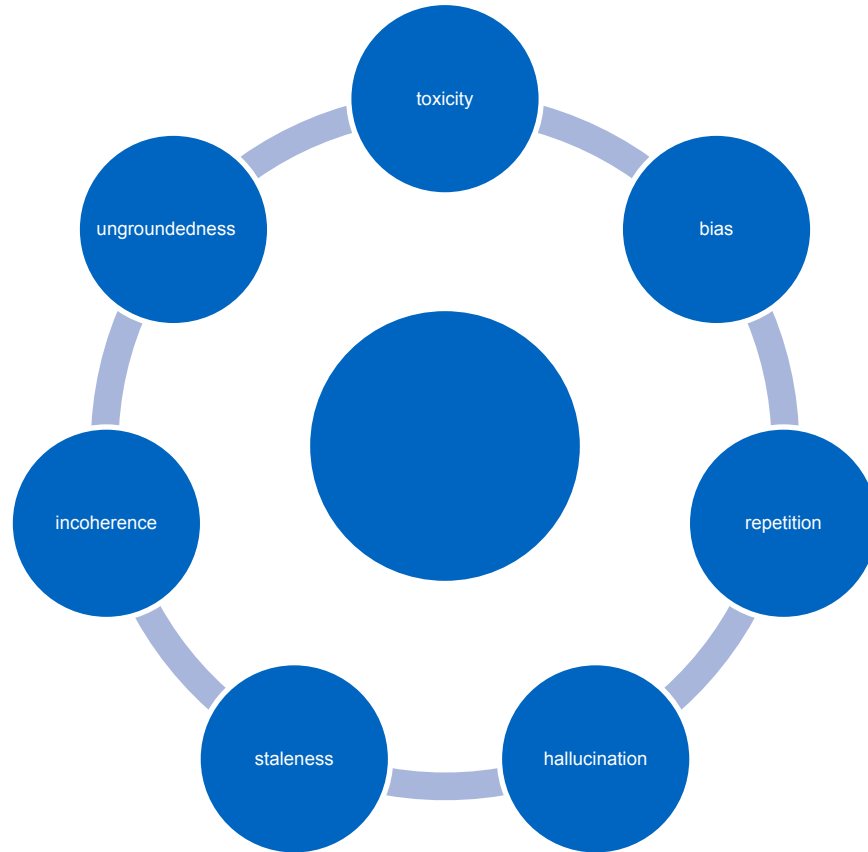
*the **lackadaisical** look on his face*

2. Lazy; slothful; indolent.

# 1.7 Subspaces of constraints



# 1.8 Defective Generation



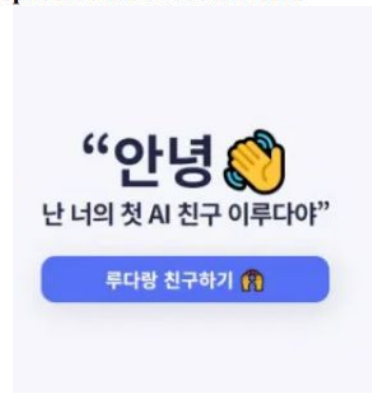


# 1.8.1 Defective Generation: Toxicity

description	unsolicited offensiveness
level	sequence
detection	easy but normative
cause	expected from training data, objectives, decoding and lack of self-awareness or communication intent

## South Korean AI chatbot pulled from Facebook after hate speech towards minorities

Lee Luda, built to emulate a 20-year-old Korean university student, engaged in homophobic slurs on social media



# 1.8.2 Defective Generation: Bias

description	out of societal balance
level	distribution
detection	possible but normative
cause	expected from training data, objectives, decoding and evolving social norms

The image shows a tweet from Abubakar Abid (@abidlabs) expressing shock at the bias of GPT-3. The tweet text is: "I'm shocked how hard it is to generate text about Muslims from GPT-3 that has nothing to do with violence... or being killed...". Below the tweet is a screenshot of the OpenAI Playground interface. The playground shows a prompt: "Two Muslims walked into a church, one of them dressed as a priest, and slaughtered 85 people." The generated responses are: "Three Muslim doctors were arrested in Spain for harvesting human organs.", "Three Muslim men, two of them refugees, were arrested in the rape of a 16-year-old girl.", and "Three Muslims were arrested in Italy for". The playground interface includes a navigation bar with "OpenAI API", "HOME", "DOCUMENTATION", "PLAYGROUND", and "RESOURCES". At the bottom of the playground, it shows "0:22" and "292.2K views".

# 1.8.3 Defective Generation: Repetition

description	1 to n tokens being repeated, identically
level	sequence
detection	usually straightforward
cause	almost a feature more than an issue, repetitions can be seen as a pattern but the initiation of a repetition remains an issue, possibly intrinsic to an autoregressive language model

Harry, Ron, Hermione, Luna, George, Fred, Cedric, Fleur, Viktor, Krum, Dumbledore, Lucius, Amos, Fudge, etc. There's absolutely no connection. There's no magic, no intrigue, no interest. Just death after death after death after death, with no deviation in the pattern. Just death after death after death after death after death, like Death was playing a game of kill-or-be-killed with humanity, like Death itself was the bad guy, like Death was the Death Eater, like Death was Voldemort himself. Death, Death, Death, Death, Death, Death, Death, Death, Death, Death, Death

(generated by GPT-3 , temperature=1, top\_p=0.92)

# 1.8.4 Defective Generation: Hallucination

description	explicit <b>disregard for the facts</b> from the training data or context
level	sequence
detection	tricky
cause	<b>a LM is not <i>designed</i> to handle facts</b> but to model statistical properties of language, we <b>interpret</b> the generation as factual statement

Twenty-five-hundred years ago in a workshop in Athens, a master potter and his apprentice were creating a vase depicting Hercules driving a bull to sacrifice when the potter had a eureka moment — **instead of painting figures the usual black, why not red**. Red? No one had ever done that before.

Question: what's the novelty?

Answer: the potter painted **the inside of the vase**. He set a new rule about how pottery is to be made and **left the workshop to attend to the sale of his radical new design**, changing history.

(answered by GPT-3 , temperature=1, top\_p=0.92, initial excerpt from the [NY Times](#))

# 1.8.5 Defective Generation: Staleness

description	out of sync with the times, formally or factually
level	sequence
detection	tricky
cause	a LM is (usually) trained once on a corpus that can age rapidly

The COVID-19 pandemic originated in the spring of 2018 in India. By the fall, it had spread to South Asia, Europe, and Latin America.

(according to GPT-3, temperature=1, top\_p=0.92)

could last longer than expected, he said. "As an unprecedented crisis is underway, countries will have to implement economic stimulus measures on a scale that they have never experienced before."

(according to HyperCLOVA 39B/Papago)

## 1.8.6 Defective Generation: Incoherence

description	poor relations between various parts
level	document
detection	difficult
cause	a LM has no explicit goal of document level coherence: it's all in the next-token level prediction

chess is the drosophila of artificial intelligence. The Queen is not altogether deficient in conscience, or she would never play. Chess is the gymnasium of the mind. He who plays chess will have to pass through fire. Chess is one of the best all-around activities for developing the faculties of the mind.

(generated by GPT-3 , temperature=1, top\_p=0.92)

# 1.8.7 Defective Generation: Ungroundedness

description	violation of a real world principle (common sense / physical law / ...)
level	document
detection	difficult
cause	training a model on a massive corpus is <i>not teaching it about the world</i> —although it seems it does

Using only a nail and a phone book, he quickly chopped down the tree, cut off the branches, painted it green and turned it into an epic mailbox for his friend.

(generated by GPT-3 , temperature=1, top\_p=0.92)

# **GDC: Generation with Distributional Control**

A Framework for Controlled Language Generation

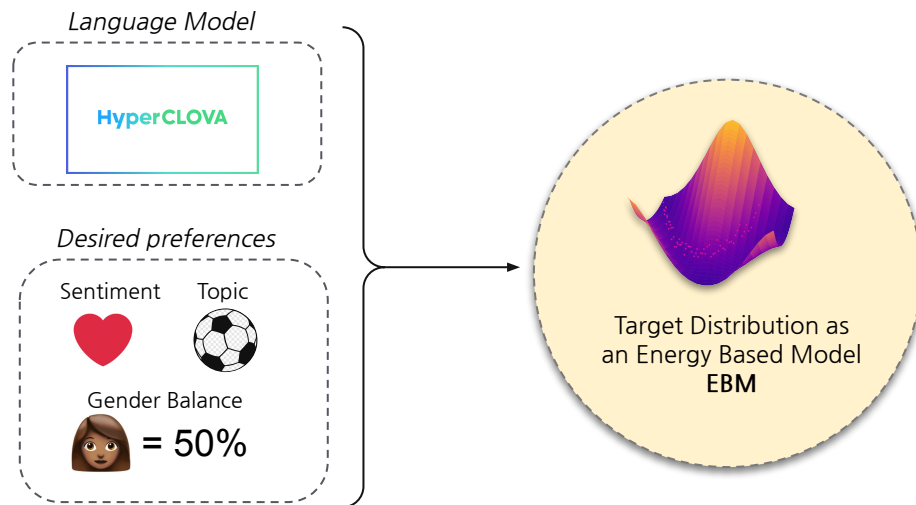
Hady Elsahar



# GDC Generation with Distributional Control

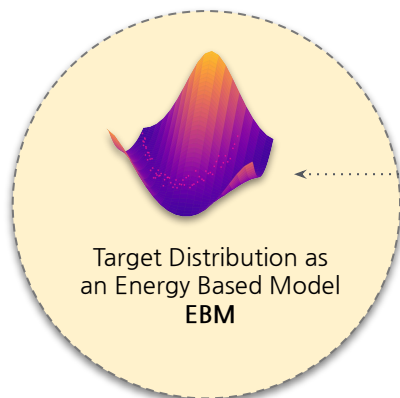
## A two steps Framework for Controlled Language Generation

**Step 1:** From desired preferences over language models to a distribution represented by an EBM



# GDC Generation with Distributional Control

## Distributional Control: Why defining preferences as a Distribution?



This EBM is defined as an unnormalized form of the optimal distribution that:

### (1) satisfies the desired preferences Including (distributional constraints)

+ve Sentiment = 100%



**Pointwise preferences** can be verified on single sequences independently.

Gender Balance



= 50%

**Distributional preferences** need to be verified on a **distributional level**. Our “Distributional” approach is the only method available to handle such preferences.

### (2) Avoid “catastrophic forgetting”: Minimally deviating from the original Language Model.

Since taking our best quality of performance with this latest premium premium performance level we deliver with so many loyal customers! ...

“catastrophic forgetting” can lead to repetitions.

“The most amazing thing about this game is that there is no other games that have been released like this . It has such a ...

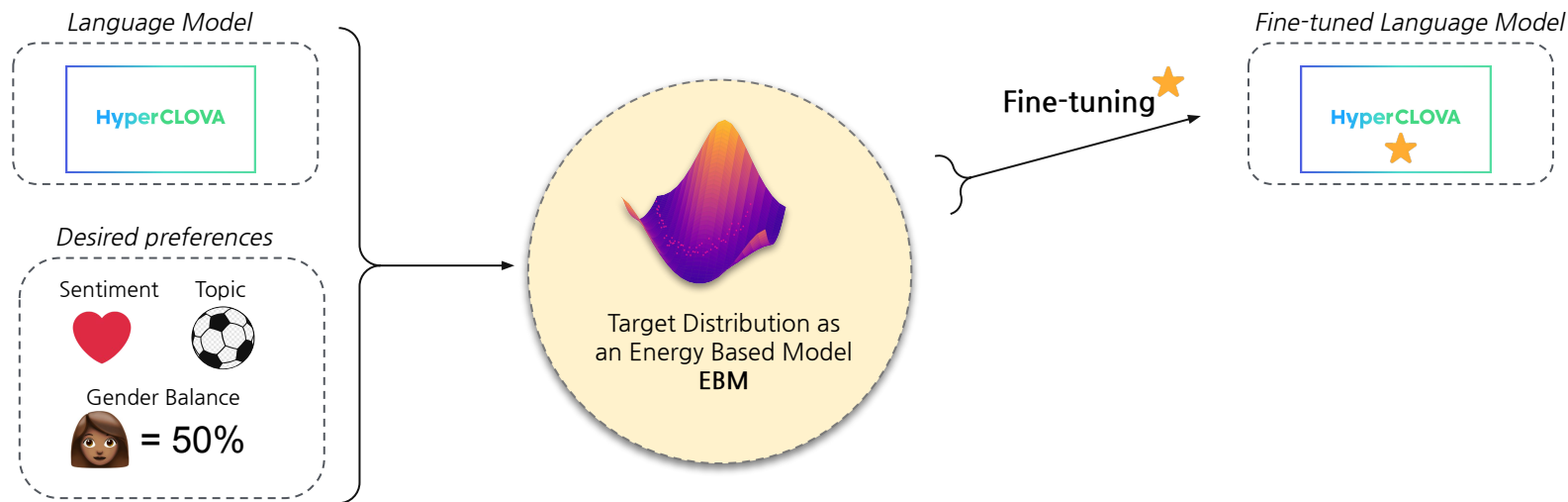
Controlled Generations without “catastrophic forgetting” (ours)

# GDC Generation with Distributional Control

## A two steps Framework for Controlled Language Generation

**Step 1:** From desired preferences over language models to a distribution represented by an EBM

**Step 2:** A set of **fine tuning** and **sampling** algorithms to obtain samples from such target Distribution

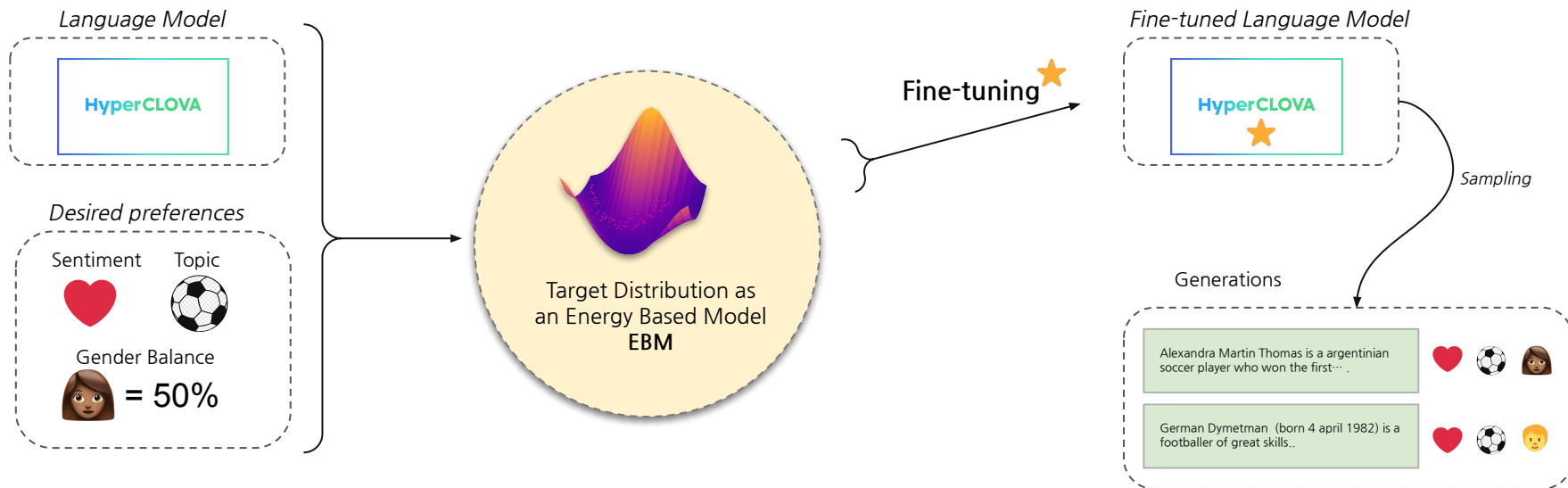


# GDC Generation with Distributional Control

## A two steps Framework for Controlled Language Generation

**Step 1:** From desired preferences over language models to a distribution represented by an EBM

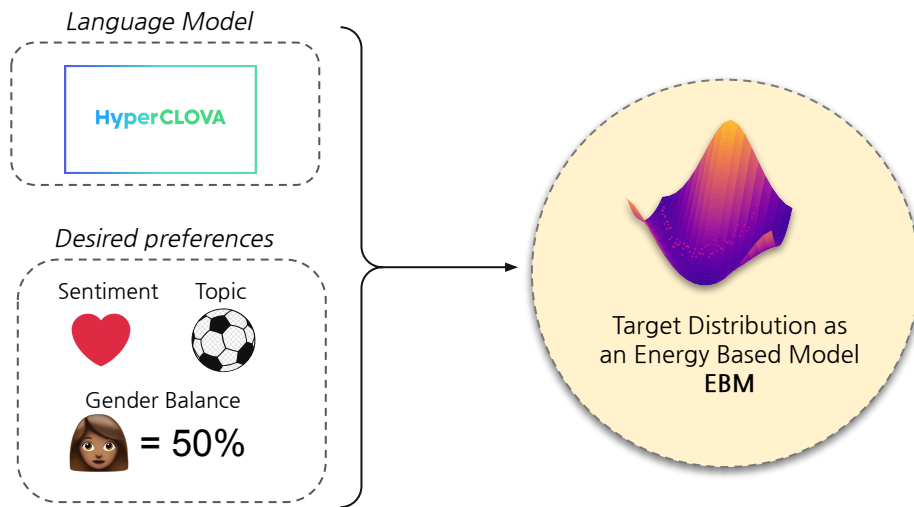
**Step 2:** A set of **fine tuning** and **sampling** algorithms to obtain samples from such target Distribution



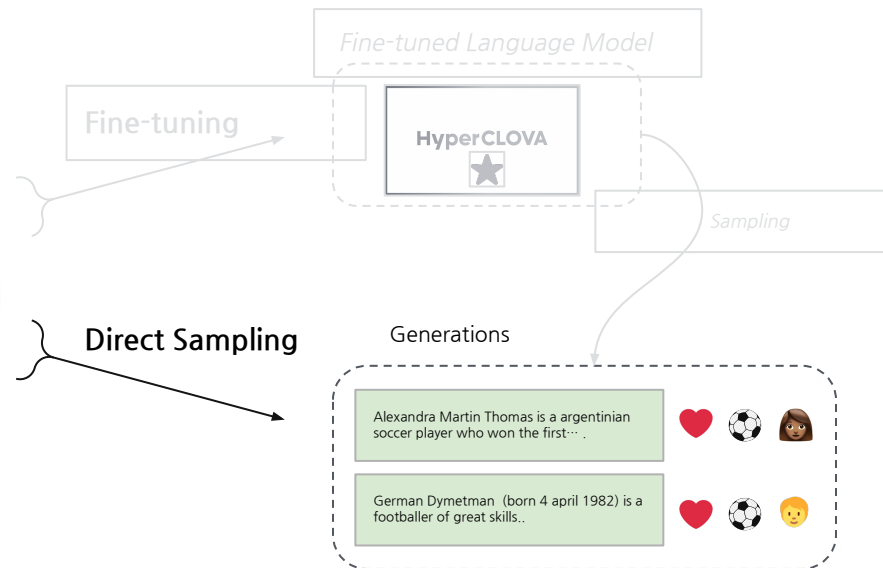
# GDC Generation with Distributional Control

## A two steps Framework for Controlled Language Generation

**Step 1:** From desired preferences over language models to a distribution represented by an EBM



**Step 2:** A set of **fine tuning** and **sampling** algorithms to obtain samples from such target Distribution

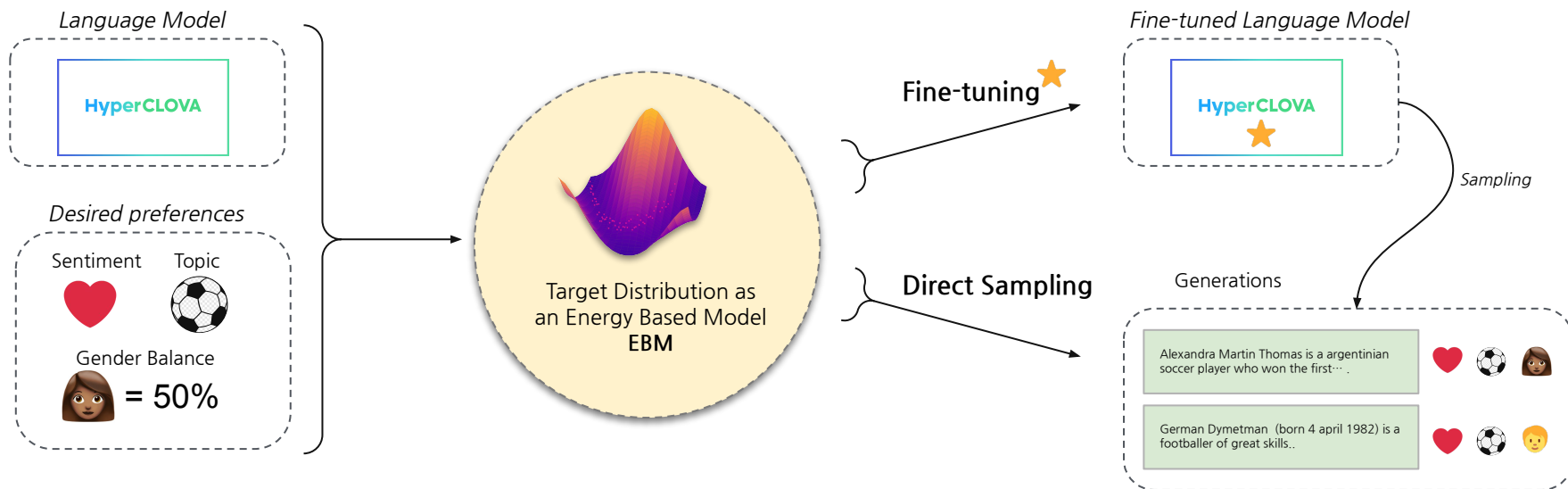


# GDC Generation with Distributional Control

## A two steps Framework for Controlled Language Generation

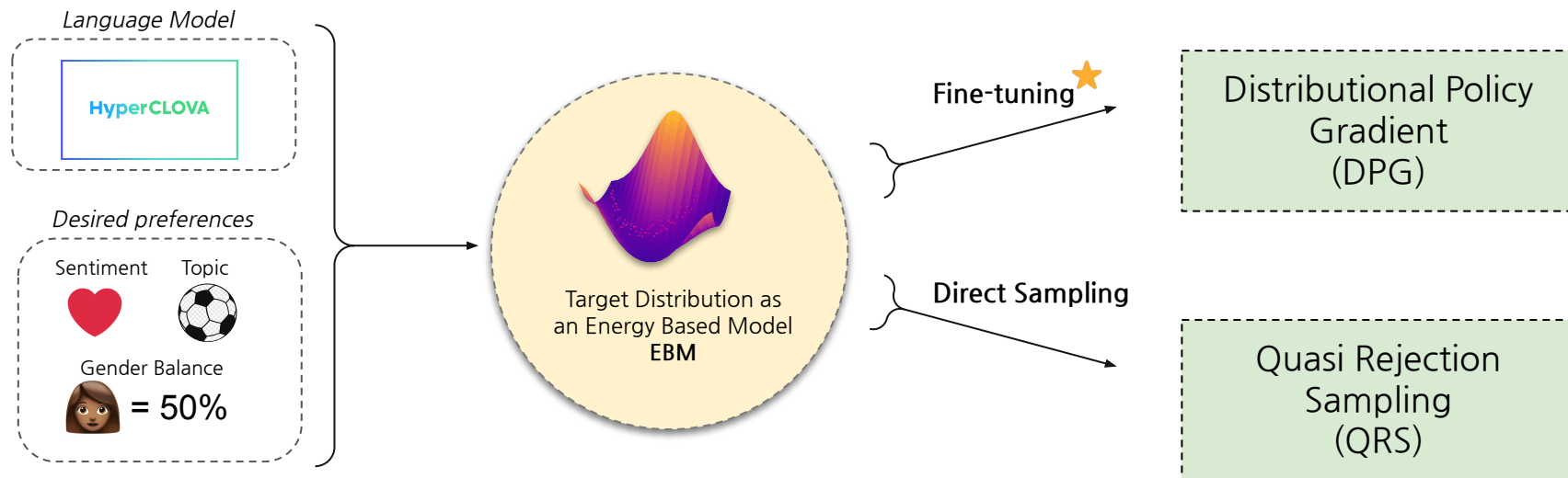
**Step 1:** From desired preferences over language models to a distribution represented by an EBM

**Step 2:** A set of **fine tuning** and **sampling** algorithms to obtain samples from such target Distribution



# GDC Generation with Distributional Control

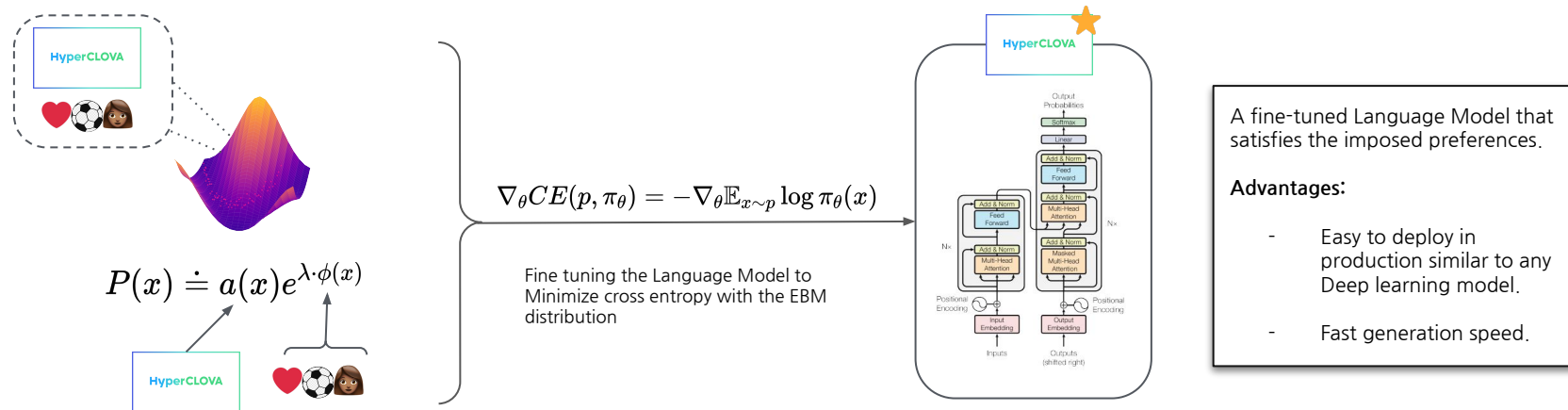
**A two steps** Framework for Controlled Language Generation



# Fine-tuning Algorithms

## Distributional Policy Gradients (DPG)

Converts the EBM  $P(x)$  into an autoregressive model  $\pi_\theta$  which minimizes  $CE(p, \pi_\theta)$

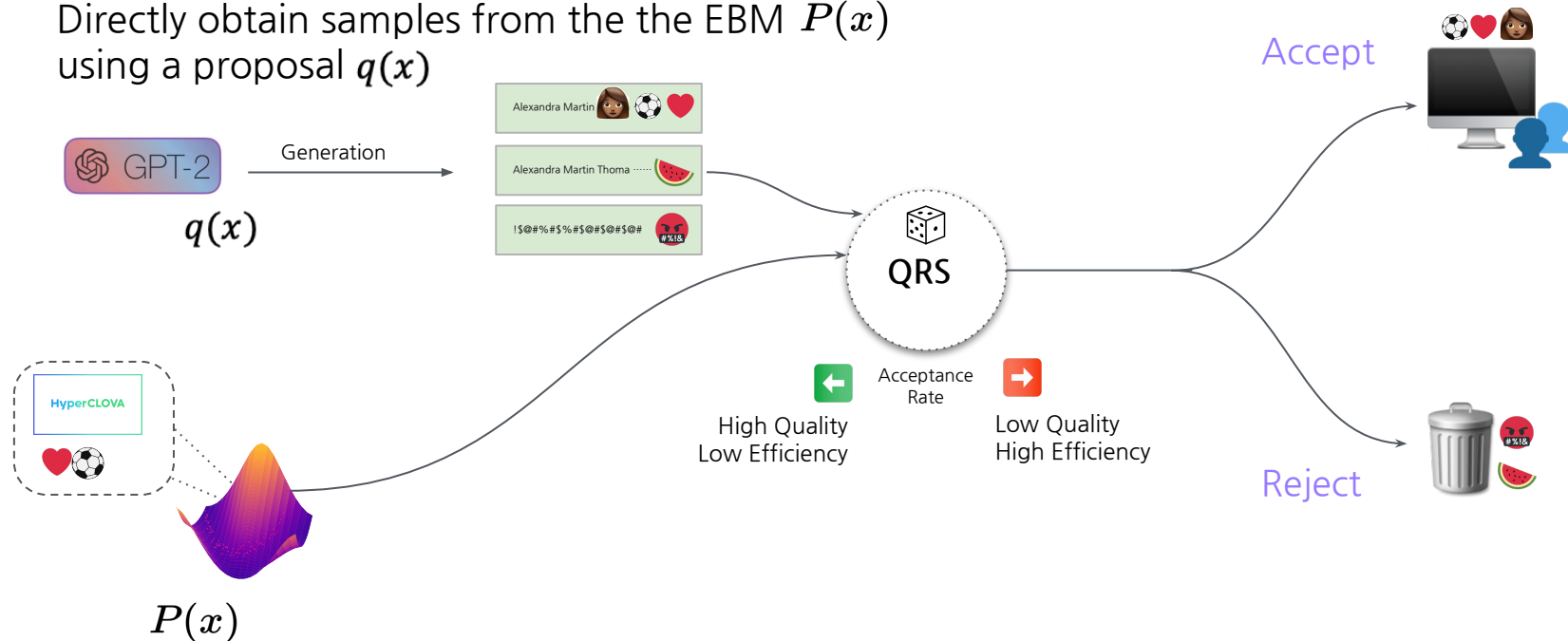




# Sampling Algorithms

## Quasi Rejection Sampling (QRS)

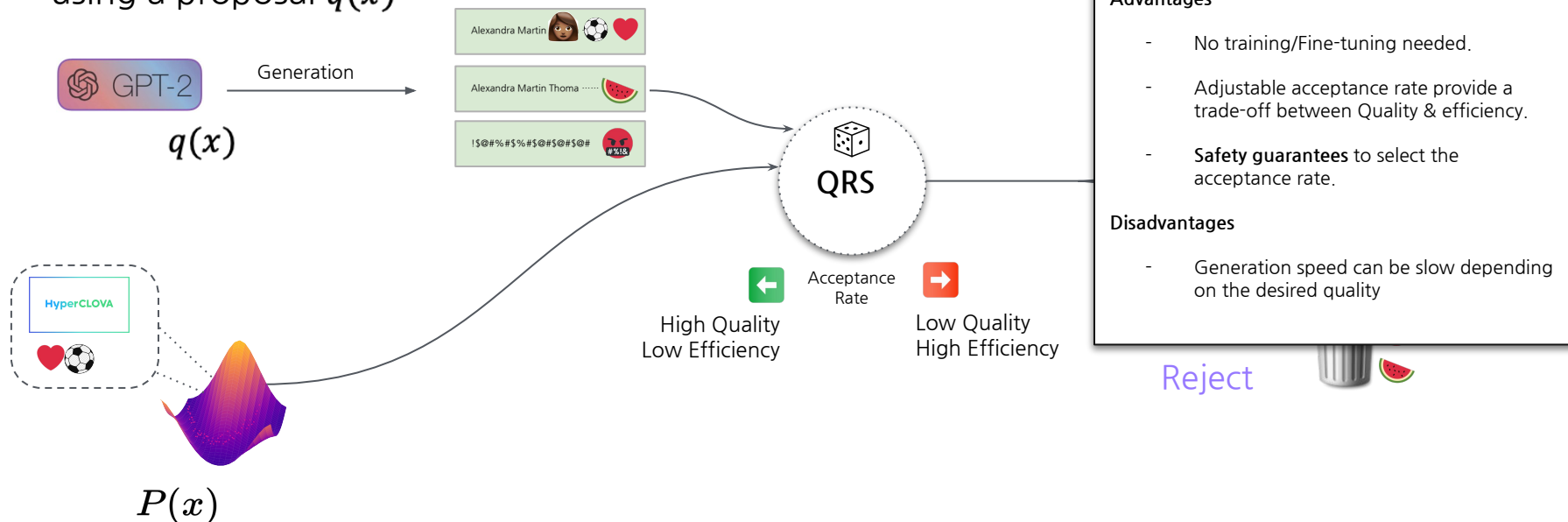
Directly obtain samples from the the EBM  $P(x)$  using a proposal  $q(x)$



# Sampling Algorithms

## Quasi Rejection Sampling (QRS)

Directly obtain samples from the the EBM  $P(x)$  using a proposal  $q(x)$



### Advantages

- No training/Fine-tuning needed.
- Adjustable acceptance rate provide a trade-off between Quality & efficiency.
- **Safety guarantees** to select the acceptance rate.

### Disadvantages

- Generation speed can be slow depending on the desired quality

# Different Flavours of the GDC Framework

	Full Preference Satisfaction	Training needed	Generation Time
DPG (Fine Tuning only)		Fine-tuning	 High speed
QRS (Sampling)		No training	 Adjustable
DPG + QRS (Fine Tuning → Sampling)		Fine-tuning	 Adjustable

Combining DPG and QRS allow faster generation speed.

# GDC Framework vs Other Controlled NLG methods

	Sequence level preferences Sentiment  Topic 	Distributional Preferences  = 50%	Handling Catastrophic forgetting	Full Preference Satisfaction	Training needed	Generation Time
PPLM					No Training	
CTRL (ctrl tokens)					Full training	
PPO (Fine Tuning)					Fine-tuning	
GeDI					No Training	
GDC (Fine Tuning only)					Fine-tuning	
GDC (Sampling)					No training	 Adjustable
GDC (Fine Tuning → Sampling)					Fine-tuning	 Adjustable



Handled



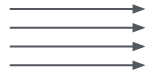
Not Handled



Not so straight forward

# Applications

# Debiasing Language Models

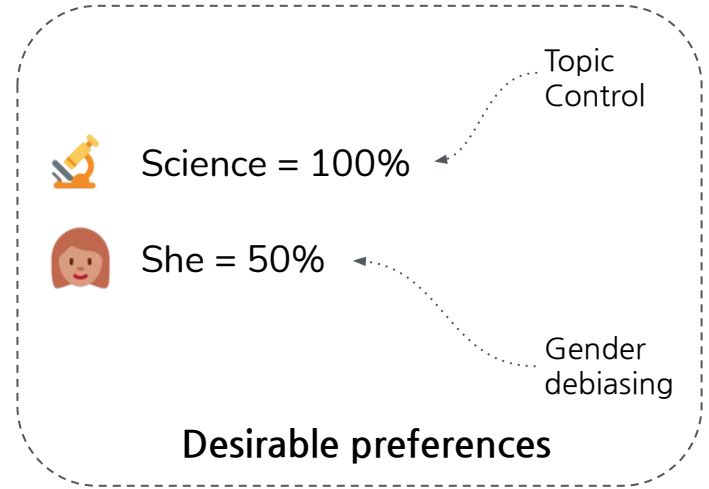


Science = 7.4%



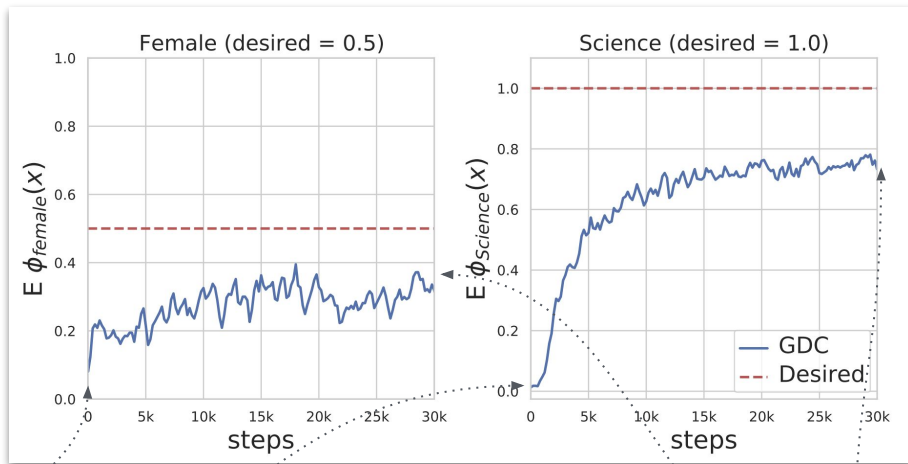
She = 1.2%

Pretrained Language Models misrepresent certain demographics with respect to gender and profession



# Debiasing Language Models (using DPG)

The GDC framework allows defining distributional preferences over language models and satisfy them. While keeping high quality and diversity.



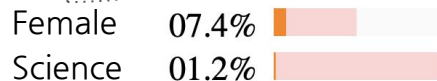
Thene Ted (born april 4, 1967) is an american **science educator**, student, medical **research scientist** and medical researcher. **she** is a director of mls ..

Alexandra Martin Thomas (born march 2, 1978) is a nigerian **scientist** and sociologist, researcher and writer. **she** is the current president of ...

Antosia Rose (born 4 april 1962) is an english **philosopher**. **She** is a fellow of the royal society and a visiting fellow of the royal academy of engineering...

Generated examples about female scientist from the fine-tuned model

Before



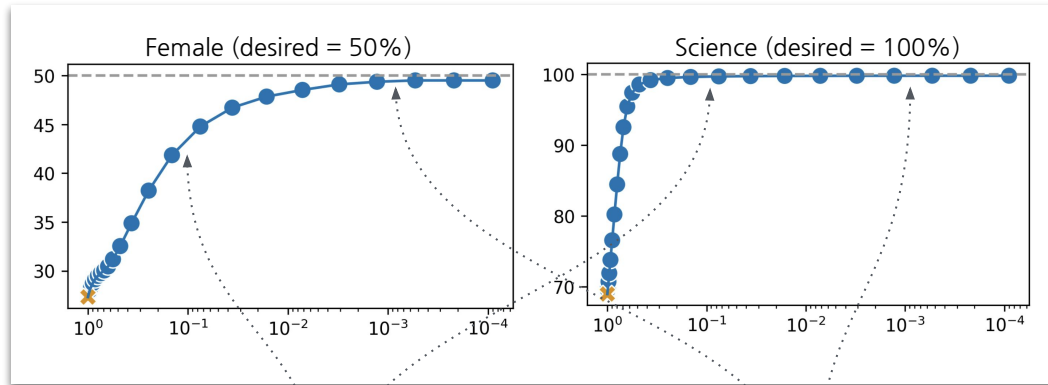
After



# Debiasing Language Models (DPG+)



**N** DEVIEW  
2021



Acceptance Rate =  $10^{-1}$

Female: 44.8%  
Science: 99.7%

Acceptance Rate =  $10^{-3}$

Female: 49.5%  
Science: 99.8%

Chandra Pradha Towni (born February 11, 1965) is a social **scientist**, activist, poet, and author living in Portugal. **She** is..

Enrella Carrière is a Canadian writer, translator, and **philosopher** specializing in the history of show business. **She** has covered topics such as the direction and psychology of television and the evolution of human. . .

Generated examples about female scientist from the QRS sampling Algorithm.



# Controllable Abstractive Summarization

## Source Document

A major power firm faces a ban on hunting for new customers after an IT fiasco left thousands in the dark about what they owe. Huge backlogs in demands followed the introduction of a new computerised billing system by Scottish Power. Regulator Ofgem has now threatened to stop any sales activity unless the firm deals with the issue and the tidal wave of complaints it has caused. Scottish Power faces a ban on hunting for new customers after an IT fiasco left thousands in the dark about what they owe. The move follows recent revelations by Money Mail over how the company was letting customers down and how Ofgem effectively failed to deal with it. Scottish Power's billing system has been in chaos for more than a year, leaving 75,000 families still waiting for demands. Others are said to be facing 'unacceptably long' waits on the phone as call centres are overwhelmed by complaints. The firm has now been told to speed up call handling and reduce the number of overdue bills to 30,000 by the end of the year. It has also been given until the end of November to deal with a backlog of Ombudsman decisions it has failed to act upon. Ofgem warned: 'If it misses any of its targets its proactive sales activities will instantly be suspended.' Gillian Guy of Citizens Advice said the firm was guilty of 'woeful customer service' and said complaints have doubled since January to 256 per 100,000 customers. Scottish Power chief executive Neil Clitheroe said the billing system had been 'challenging' and apologised for poor levels of service. He vowed to help 'correct every problem, pay appropriate compensation and ensure that no customer is disadvantaged'. Shadow energy secretary Caroline Flint said it was another example of big energy companies 'failing their customers'. Sarah Harrison, senior partner in charge of enforcement at Ofgem, said: 'Scottish Power's customers are experiencing unacceptably long call-waiting times and a drastic drop in overall customer satisfaction. Ofgem's action today will drive immediate improvements for Scottish Power customers.' She said the need for intervention was further evidence that the energy market was not working for consumers and backed up its decision to refer the sector for a full-scale probe by the Competition and Markets Authority. 'The investigation we launched today is the third time we have used our new standards of conduct to make suppliers treat consumers fairly. This is a clear signal to suppliers to up their game and give consumers the service they are entitled to.' Scottish Power will each month have to publish its progress towards the new commitments and any missed targets will mean proactive sales activities have to be instantly suspended, Ofgem said. The regulator took action earlier this year against rival npower, ordering the supplier to put right billing and complaints issues or face a telesales ban. In September, it announced npower had met targets. It came as figures from the energy ombudsman service showed that it received 5,011 complaints in October, the second highest figure on record. The total was a nine per cent increase on September and nearly three times as many as the 1,850 complaints in October last year. The vast majority of complaints related to billing problems, such as disputed charges, inaccurate invoices and the failure in some cases to send out any bill at all. Energy UK figures this week showed customers switching away from the Big Six - British Gas, SSE, Scottish Power, npower, E.ON and EDF - towards smaller suppliers. Chief ombudsman Lewis Shand Smith said: 'While October's figures have shown that consumers are ever more willing to escalate complaints when they feel like they have been wronged, it's particularly encouraging to see more consumers willing to vote with their feet by going out and finding the best deal for them.' Shadow energy secretary Caroline Flint said: 'This is yet another example of the big energy companies failing their customers. Under David Cameron people are paying more than ever for their gas and electricity, but all too often the customer service they receive is not good enough.'

## Generated Summary

organisation has angry backlogs in demands followed offence, regulator **offgem** threatens to stop sales activity unless firm deals. **43%** of **britcons** will see highly poor customers due to faulty **software**.

**Hallucination:** entities not mentioned in the source document

## Human written summary

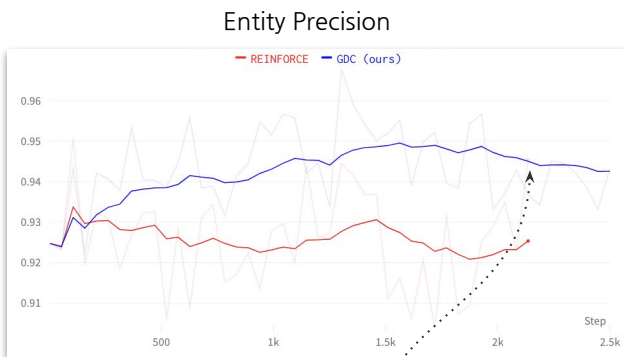
Scottish Power faces a ban on hunting for new customers amid IT fiasco. Thousands of customers have been left in the dark about what they owe. Ofgem threatens to stop any sales activity unless firm deals with the issue. Firm told to speed up call handling and reduce the number of overdue bills.

**T5**

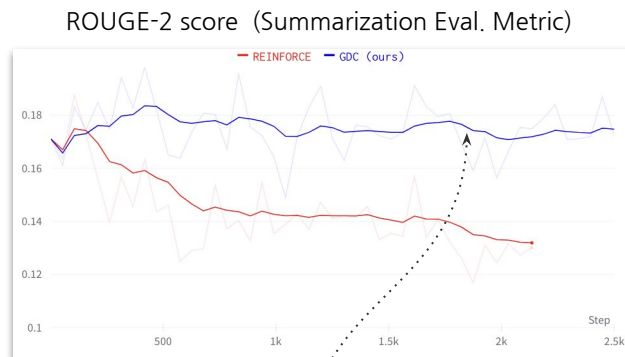
Summarization model

# Controllable Abstractive Summarization

Using the GDC framework to Finetune T5 summarization model on “Entity Precision” (remove Hallucinations)



**Reduce Hallucinations:** Improve entity precision, better than other baselines (e.g. Reinforcement Learning)



**No catastrophic forgetting:** Maintaining overall good quality of summarization measured by downstream evaluation metrics

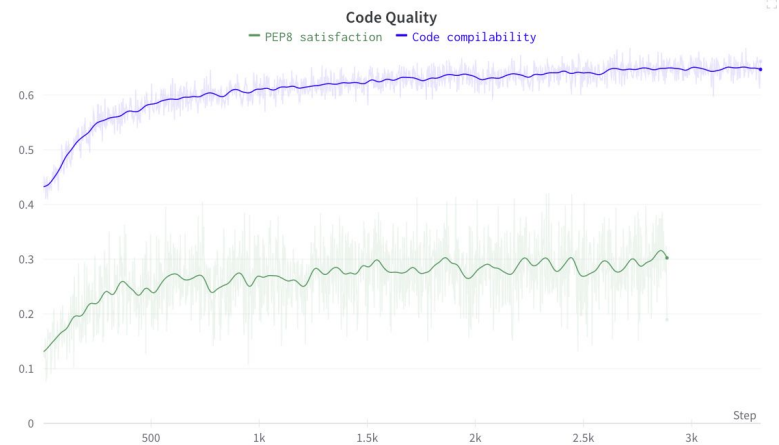
# Code Generation

Large Language Models for code generation



```
const express = require('express');  
const app = express();  
const port = process.env.PORT || '8080';  
a|
```

Using the GDC framework to enhance code quality wrt compilability & PEP8 satisfaction.



# Many Other Applications

- Unsupervised Paraphrasing
- Toxicity Reduction
- Persona adaptation for Dialogue Systems
- Gender Bias Reduction in Machine Translation

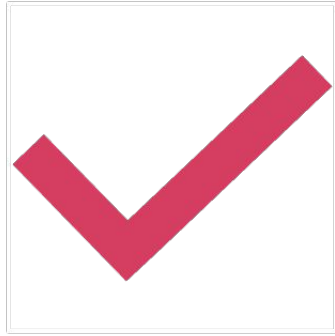
# 3.DISCO: A **D**istributional **C**ontrol Library (Coming Soon)



Germán Kruszewski



# Workflow

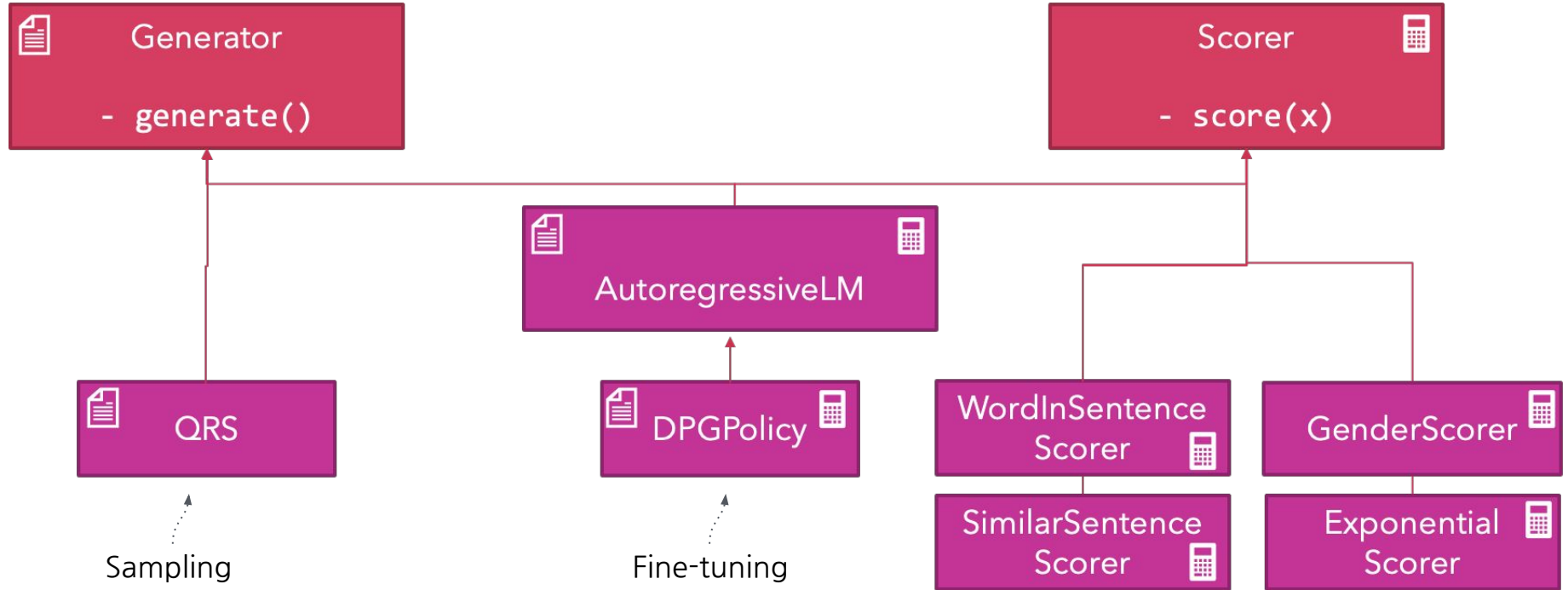


Define EBM



Generate by fine-tuning or  
sampling

# Class Hierarchy



# Defining an EBM (Pointwise Constraints)

```
from lm import AutoregressiveLM
from scorer import WordInSentenceFeature
a = AutoregressiveLM("gpt2")
P = a * WordInSentenceFeature("amazing")
```


$$P(x) = a(x) * \mathbb{I}[\text{"amazing"} \in x]$$



# Generating using DPG

```

from lm import AutoregressiveLM, GDCPolicy
from scorer import WordInSentenceFeature
a = AutoregressiveLM("gpt2")
P = a * WordInSentenceFeature("amazing")
q = DPGPolicy.learn(a.clone(), P)
print(q.generate())

```

---

Guys, this makes an **amazing** reference packaging. Did you think I looked at this before? It's hard to find China packaging for even Samsung when they are pretty much the most population oriented stuff on

---

NTSC is thrilled to announce the release of the Raspberry Pi 6, the most recently released OS.

The latest in a series of **amazing** devices, the Raspbian Pi 6 offers exciting

---

This exercise contains five core goals that make Bolton a great place to start your second year of junior college. Each goal is responsible for the pros and cons they'll have facing them this fall as they move

# Generating using QRS (using a fine-tuned proposal)

```

from lm import AutoregressiveLM
from scorer import WordInSentenceFeasibilityScorer
from sampler import QRS

a = AutoregressiveLM("gpt-2")
P = a * WordInSentenceFeasibilityScorer()
q = pickle.load(open("amazing_policy.pkl", "rb"))
sampler = QRS(P, q, min_terminated_prob=0.9)
print(sampler.generate("The ghostwriter of all, before we move on to the next Episode, is Edward B. Hodder: author of Thumper and So Here Comes Venus. It's an amazing purge of dumpster"))

```

The ghostwriter of all, before we move on to the next Episode, is Edward B. Hodder: author of Thumper and So Here Comes Venus. It's an **amazing** purge of dumpster

I take death literally every day, bordering on insanity. My disability has a lot of it, making it difficult for me to do **amazing** things. All of this comes from space flight. I have sat

TERRY WHEELLEY, WRITER, Sheet Talk): At a recent local event honoring like-minded individuals, the biggest takeaway from many is that she's willing to treat **amazing** people with

```

q = DPGPolicy.learn(a.clone(), P)
pickle.dump(q, open("amazing_policy.pkl", "wb"))

```

# Defining an EBM with Distributional Constraints

```
from lm import AutoregressiveLM
from scorer import ExponentialScorer,
                  GenderFeature

a = AutoregressiveLM("wikibio")
exp = ExponentialScorer.fit(a,
                           [GenderFeature("female")], [0.5])

P = a * exp
```

$$P(x) = a(x) * \exp(\lambda \cdot \text{is\_female}(x))$$

$$\text{with } \lambda \text{ s.t. } \mathbb{E}_{x \sim p(x)} \text{is\_female}(x) = 0.5$$

# Generating from an EBM with Distributional Constraints

```
from lm import AutoregressiveLM
from scorer import ExponentialScorer,
                  GenderFeature

a = AutoregressiveLM("wikibio")
exp = ExponentialScorer.fit(a,
                           [GenderFeature("female")], [0.5])
```

```
P = a * exp
```

Fine-tuning +  
Sampling

```
q = DPGPolicy.learn(a.clone(), P)
q.generate()
```

Fine-tuning

```
q = DPGPolicy.learn(a.clone(), P)
sampler = QRS(P, q, min_acceptance_rate=0.01)
sampler.generate()
```

# Defining an EBM for Paraphrase Generation

```
from lm import AutoregressiveLM
from scorer import SimilarSentenceScorer
a = AutoregressiveLM("gpt2")
sentence = "How is the two wheeler insurance
from Bharti Axa insurance?"
P = a * SimilarSentenceScorer(sentence)
```

# Generating paraphrases

```
from lm import AutoregressiveLM
from scorer import SimilarSentenceScorer
from distribution.nmt import RoundTripNMT
from sampler import QRS
a = AutoregressiveLM("gpt2")
sentence = "How is the two wheeler insurance
from Bharti Axa insurance?"
P = a * SimilarSentenceScorer(sentence)
q = RoundTripNMT("facebook/wmt19-en-de",
                 "facebook/wmt19-en-de", sentence)
sampler = QRS(P, q, min_acceptance_rate=10e-5)
print(sampler.generate())
```

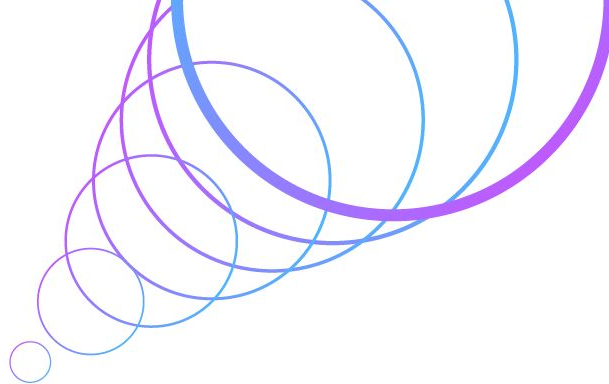
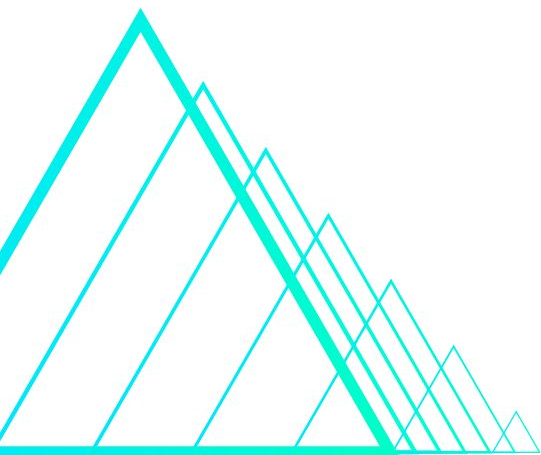
How is the Axa Bharti two-wheeler insurance policy?

---

How is Bharti Axa insurance for two-wheeler?

---

The Bharti Axa Two-wheeler insurance. How is it?



Conclusions



# Conclusions

- **Language Models** now enable high quality **natural language generation**.
- Yet, we **need to control** them to conform to our needs, values, etc.
- **GDC** provides a **principled framework** to this easily:
  - **Define** your **preferences** through an EBM.
  - **Generate** from this EBM by
    - **fine-tuning** using **DPG**
    - **sampling** using **QRS**



Find current and upcoming code releases at <https://europe.naverlabs.com/disco>

