



# КМУ

## Большие языковые модели для работы с векторной графикой

**Малашенко Б.Т, ИТМО, ФИТИП**

---

Ефимова В.А, к.т.н.

# Векторная графика: где применяется?

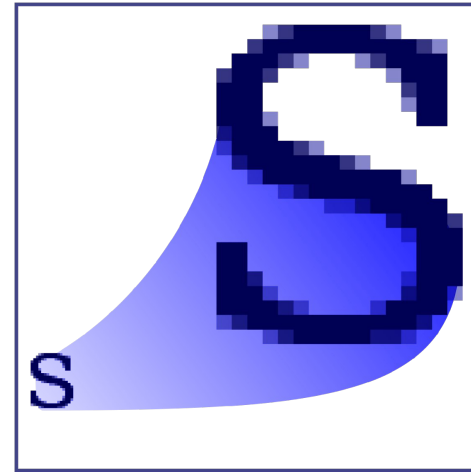
- Векторные изображения применяются для логотипов, иллюстраций, иконок, шрифтов и т. д.
- Главное преимущество — масштабирование без потери качества
- Особенность — простота и абстрактность формы.



# Векторная графика

Векторная графика (SVG) описывает изображения с помощью формул, а растровая — с помощью пикселей.

- Фигуры формируют изображение;
- Каждая фигура задаётся математически;
- Для каждой фигуры можно задать цветовые характеристики.



**Raster**  
GIF, JPEG, PNG



**Vector**  
SVG

# Использования с БЯМ

Большие языковые модели (БЯМ) — это нейросети, обученные обрабатывать и генерировать текст на естественном языке.

Поскольку SVG — это текст в формате XML, БЯМ могут читать, интерпретировать и генерировать векторную графику так же, как и обычный текст.



```
<svg>  
width = 100  
height = 100  
<path >  
...  
</svg>
```

# Цель и задачи

Цель – исследовать возможности применения БЯМ для генерации, редактирования и анализа векторной графики.

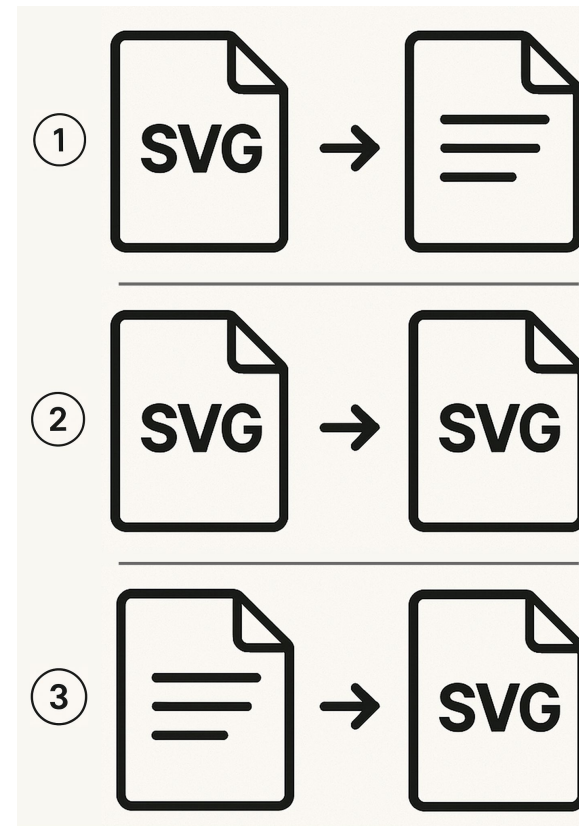
Для этого ставятся следующие задачи:

1. анализ существующих методов оценки БЯМ при работе с SVG-графикой;
2. сравнительная оценка различных моделей;
3. разработка универсальной методики оценки БЯМ для векторной графики.



# Три основных направления при работе с SVG

1. Генерация SVG на основе текстового запроса.
2. Редактирование SVG по текстовой инструкции.
3. Распознавание деталей векторного изображения.



# Как оценивать БЯМ?

Бенчмарк — это стандартизированный набор задач и метрик, предназначенный для объективной оценки качества работы модели.

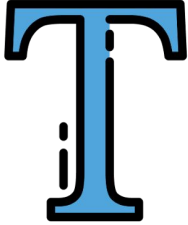

Основные бенчмарки для работы с векторной графикой:


- Image-text bridging;
- SVGEEditBench;
- SVG Taxonomy;
- VGBench (VGen + VGQA);
- SGP-Bench.

# VG Bench

VGQA: Оценка понимания моделью цветов, форм и действий.

VGen: Генерация SVG по запросу.

VGQA	
Question:	Question:
	
What is the dominant color of the text icon in the SVG image?	What does this SVG image primarily represent?
Answer:	Answer:
A. Green B. Yellow <u>C. Blue</u> D. Red	A. A mobile phone <u>B. A playing card</u> C. A book D. A photo frame

VGen
Question:

What emotion or concept is this SVG image primarily representing?
Answer:
A. Freedom <u>B. Love or Romance</u> C. Migration D. Solitude

VGen
Caption:
The image is a simple black and white silhouette of a bottle with a triangular label. The bottle has a narrow neck and a wider base, resembling a typical glass bottle shape. The label has a prominent downward-pointing arrow in the center. The silhouette is stylized and lacks detail, making it suitable for a logo or an icon representing a bottle.


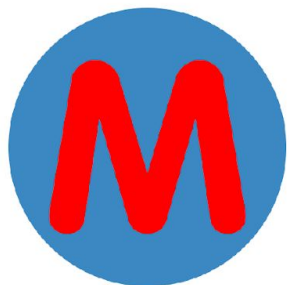




Ground truth image:





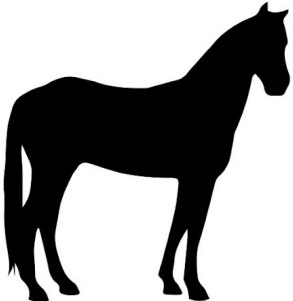



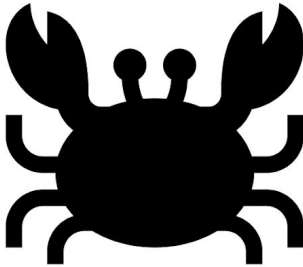
# SVGEEditBench

Шесть простых задач по редактированию SVG: замена цвета, установка контура, переворот, прозрачность, обрезка, сжатие.

Query: The following code is the SVG code for the emoji 'circled latin capital letter m'.	Task [change color]: Please generate an SVG code that changes the part of the emoji with a #FFF color to red.	Task [set contour]: Please generate an SVG code that draws a black line around the part of the emoji with a #FFF color.	Task [upside down]: Please flip this emoji upside down.	Task [transparency]: Please make this emoji transparent by half.	Task [crop to half]: Please trim the right half and keep the left half.
Query SVG:	Ground truth:	Ground truth:	Ground truth:	Ground truth:	Ground truth:
					

# SGP-Bench

Understanding, different categories.

<p>Question:</p>  <p>What type of animal is represented by the object in the image?</p>	<p>Question:</p>  <p>What object is depicted in the image?</p>	<p>Question:</p>  <p>What is the person in the image doing?</p>	<p>Question:</p>  <p>What is the color of the object in the image?</p>	<p>Question:</p>  <p>How many legs does the object have?</p>
<p>Answer:</p> <p>A. <u>Horse</u> B. Cat C. Dog D. Elephant</p>	<p>Answer:</p> <p>A. Book B. <u>Newspaper</u> C. Notebook D. Magazine</p>	<p>Answer:</p> <p>A. Running B. <u>Walking</u> C. Sitting D. Jumping</p>	<p>Answer:</p> <p>A. Blue B. Green C. Yellow D. <u>Pink</u></p>	<p>Answer:</p> <p>A. 4 B. 6 C. 10 D. <u>8</u></p>

# Агрегация бенчмарков

Нормализуем метрики каждого бенчмарка, затем посчитаем усредненные значения.

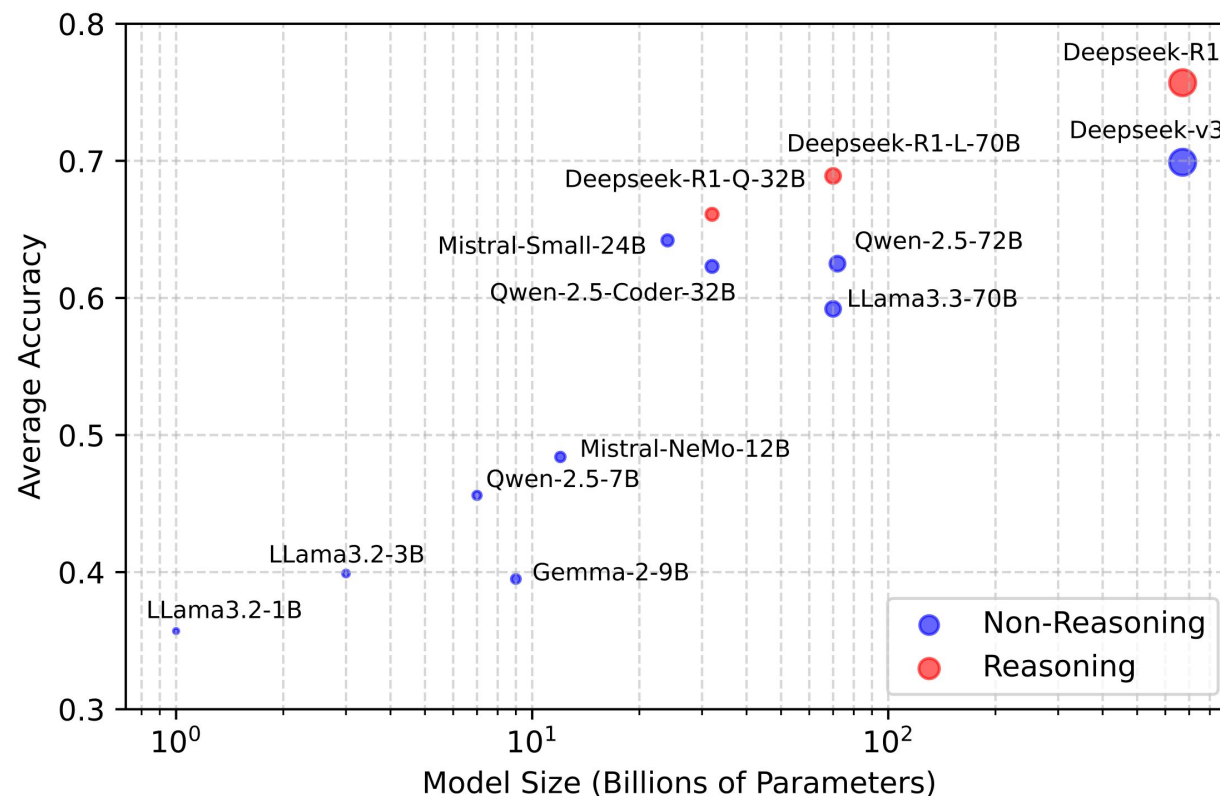
Результаты работы моделей.

	Model	Generation	Editing	Understanding		Avg
		VGen	SVGEdit	VGQA	SGP	
A	IconShop	0.816	-	-	-	-
	Deepseek-v3	0.923	0.642	0.460	0.606	0.699
	Gemma-2-9B	0.857	0.084	0.060	0.429	0.395
	LLama3.2-1B	0.805	0.073	0.220	0.168	0.357
	LLama3.2-3B	0.834	0.063	0.193	0.406	0.399
B	LLama3.3-70B	0.863	0.460	0.347	0.558	0.592
	Mistral-NeMo-12B	0.848	0.267	0.253	0.419	0.484
	Mistral-Small-24B	0.866	0.722	0.187	0.491	0.642
	Qwen-2.5-7B	0.865	0.081	0.367	0.475	0.456
	Qwen-2.5-72B	0.872	0.533	0.372	0.567	0.625
C	Qwen-2.5-Coder-32B	0.890	0.545	0.333	0.532	0.623
	Codestral-2501	0.889	<u>0.901</u>	0.273	0.484	0.723
	Deepseek-R1	<u>0.957</u>	0.666	<u>0.587</u>	<u>0.710</u>	<u>0.757</u>
D	Deepseek-R1-Q-1.5B	∅	0.109	0.067	0.134	-
	Deepseek-R1-Q-32B	0.858	0.721	0.407	0.401	0.661
	Deepseek-R1-L-70B	0.861	0.751	0.387	0.521	0.689
E	Qwen-Turbo	0.858	0.574	0.296	0.477	0.606
	Qwen-Max	0.906	0.599	0.440	0.528	0.663

# Как влияет reasoning?

Reasoning-модели и reasoning-based дистилляты показывают лучшее среднее качество в различных размерах.

Сравнение обычных и reasoning БЯМ.



# Результаты

По итогам этой работы:

- рассмотрены бенчмарки для разносторонней оценки БЯМ;
- суммарно приведены оценки для 21 модели;
- предложена агрегирующая оценка, оценивающая качество модели в генерации, редактировании и распознавании SVG.

# Заключение

Векторная графика в сочетании с большими языковыми моделями открывает путь к более доступным инструментам.

Для продвижения в этом направлении важно:

1. разрабатывать универсальные методы оценки качества работы моделей;
2. расширять и улучшать датасеты;
3. изучать новые методы обучения БЯМ;
4. создавать удобные пользовательские интерфейсы.



**СПАСИБО ЗА ВНИМАНИЕ!**

**IT's MO** *re than a*  
**UNIVERSITY**

Санкт-Петербург, 2025