Learning with Real and Unreal Data in the Era of Foundation Models

Rogerio Feris
Principal Scientist and Manager
MIT-IBM Watson AI Lab, IBM Research

November 08, 2023

30 Years Ago …

RETRATO FALADO

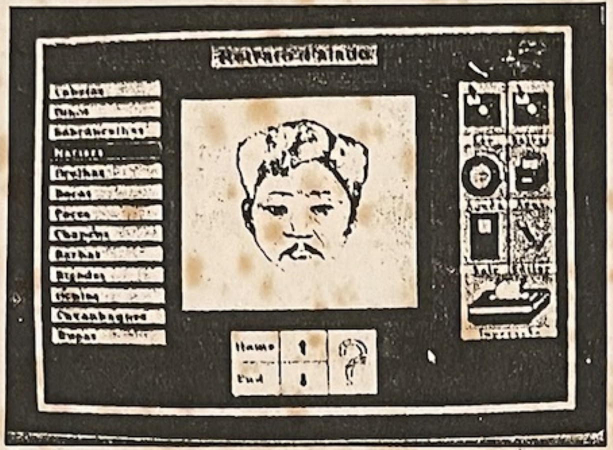
Computador vai desenhar rosto de suspeitos

□ Três jovens criaram um programa que garante maior rapidez e precisão no momento de identificar pessoas acusadas de crimes

EUNICE LEME
Correspondente

Rio Grande — Os alunos do curso de Processamento de Dados do Colégio Técnico Industrial (CTI) de Rio Grande estão desenvolvendo um programa de computação que permite a criação do retrato falado de uma pessoa. O programa poderá em breve ajudar a polícia a identificar criminosos com maior rapidez e precisão. O sistema está em fase de aperfeiçoamento e a previsão é de que no próximo ano já esteja adequado às necessidades da Polícia Civil de Porto Alegre. O *software* é inédito no país. O que existe até hoje são sistemas importados, segundo afirma Delnir Monteiro Lemos, professor do curso. Em Brasília, a Polícia Federal também estuda a adoção do programa criado em Rio Grande. Se for aprovado, a PF pode utilizá-lo em seus escritórios de todo o país.

Os programas, que de[...] computador o rosto de u[...] com as mesmas técnicas[...] je pela polícia, embora c[...] dade e a precisão da informática, foram criados por três alunos adolescentes do curso de Processamento do CTI. Alessandro Bicho, de 17 anos, Marco Antônio Gomes, 18 anos, e Rogério Féris, de 18 anos, formaram-se no início do ano e vão continuar trabalhando no projeto por mais seis meses. Os primeiros resultados foram animadores. Apresentado como trabalho de conclusão de curso em várias feiras e exposições do Estado, o progra-



Na tela: *a figura de uma pessoa fica mais precisa*

José Doval/ZH

Feito à mão: *Paulo Severo usa um kit para identificar*

Cláudia Zunino/ZH

Criativos: *Marco (de óculos), Alessandro (sentado) e Rogério querem ver o projeto usado em investigações*

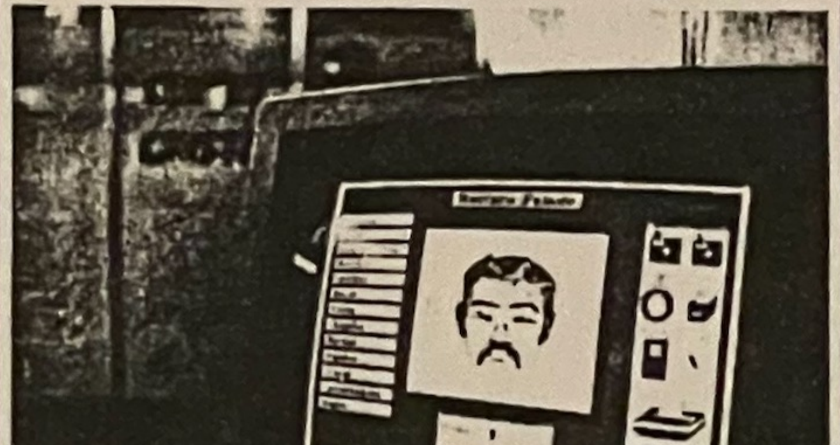
O projetor de imagens é hoje a principal arma

O uso do computador no desenho de retratos falados reduz eta[...]balho e torna as informa[...] precisas, admite o dese[...]crivão Paulo Severo Nu[...]ealiza à mão este tipo de [...]ra toda a Polícia Civil gaúcha. Na elaboração de retratos falados, Severo utiliza um retroprojetor, um livro mostruário e um *kit* composto por lâminas de projeção, com figuras de cabelos, sobrancelhas, olhos, narizes, lábios, barbas, bigodes, orelhas e formatos de rostos variados. O livro é mostrado às testemunhas que vão selecionando os traços que mais se assemelham ao acusado de um crime. Com a superposição das lâminas, o desenhista forma uma figura, que, projetada numa parede, é reproduzida numa folha de papel, através de um desenho.

No programa de computador, explica, a imagem criada a partir da

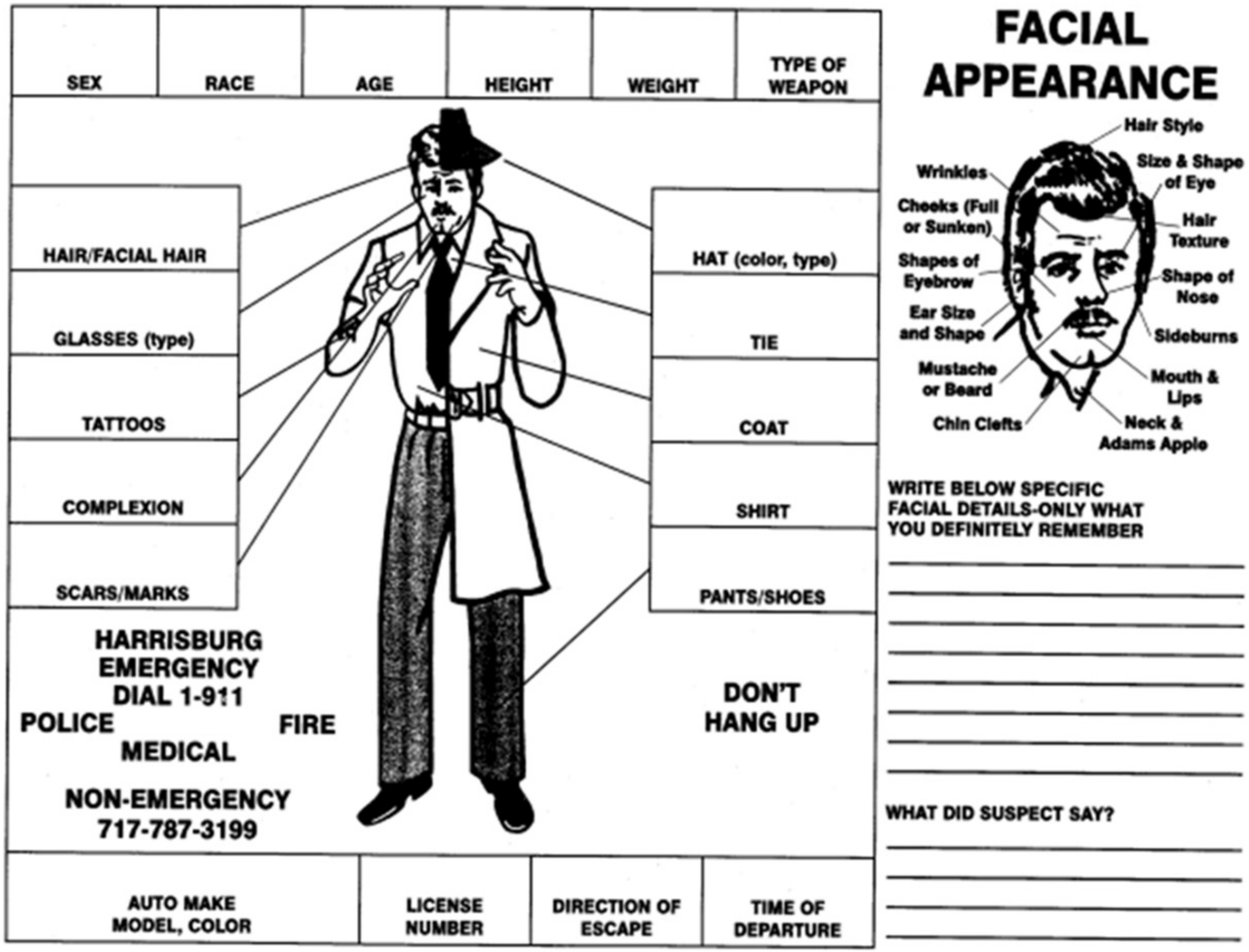
Retrato falado é feito com informações das testemunhas

O programa possui 600 imagens de cada detalhe do rosto



IBM Intelligent Video Analytics (2013)

People Search based on a Suspect Description Form



TECH & MEDIA

Facial recognition's 'dirty little secret': Millions of online photos scraped without consent

People's faces are being used without their permission, in order to power technology that could eventually be used to surveil them, legal experts say.

The New York Times

Facial Recognition Is Accurate, if You're a White Guy

Share full article

By Steve Lohr

Feb. 9, 2018

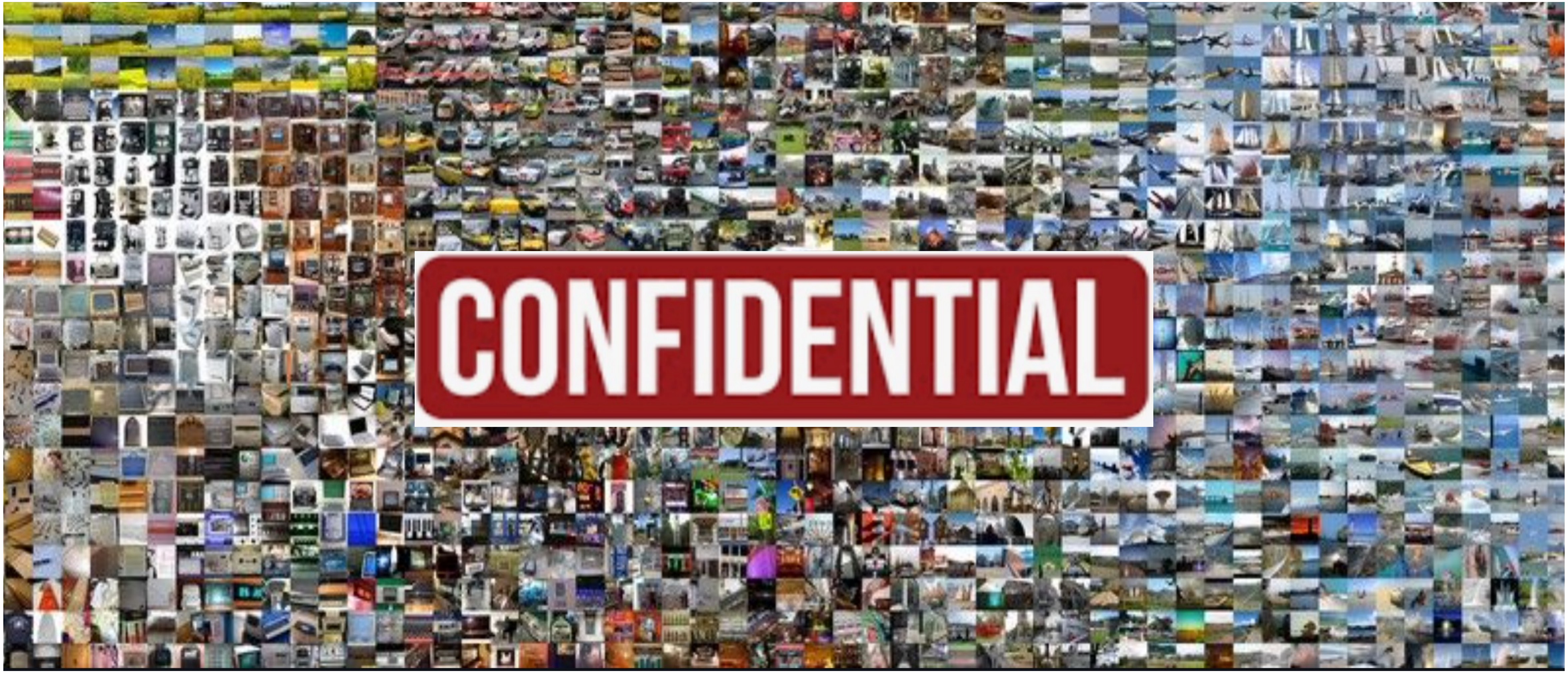
Facial recognition technology is improving by leaps and bounds. Some commercial software can now tell the gender of a person in a photograph.

When the person in the photo is a white man, the software is right 99 percent of the time.

But the darker the skin, the more errors arise — up to nearly 35 percent for images of darker skinned women, according to a new study that breaks fresh ground by measuring how the technology works on people of different races and gender.

Era of Foundation Models: Pre-training on Massive Datasets of Web Images and Videos

Google JFT -- 3Billion images



Promising work-around: synthetic data

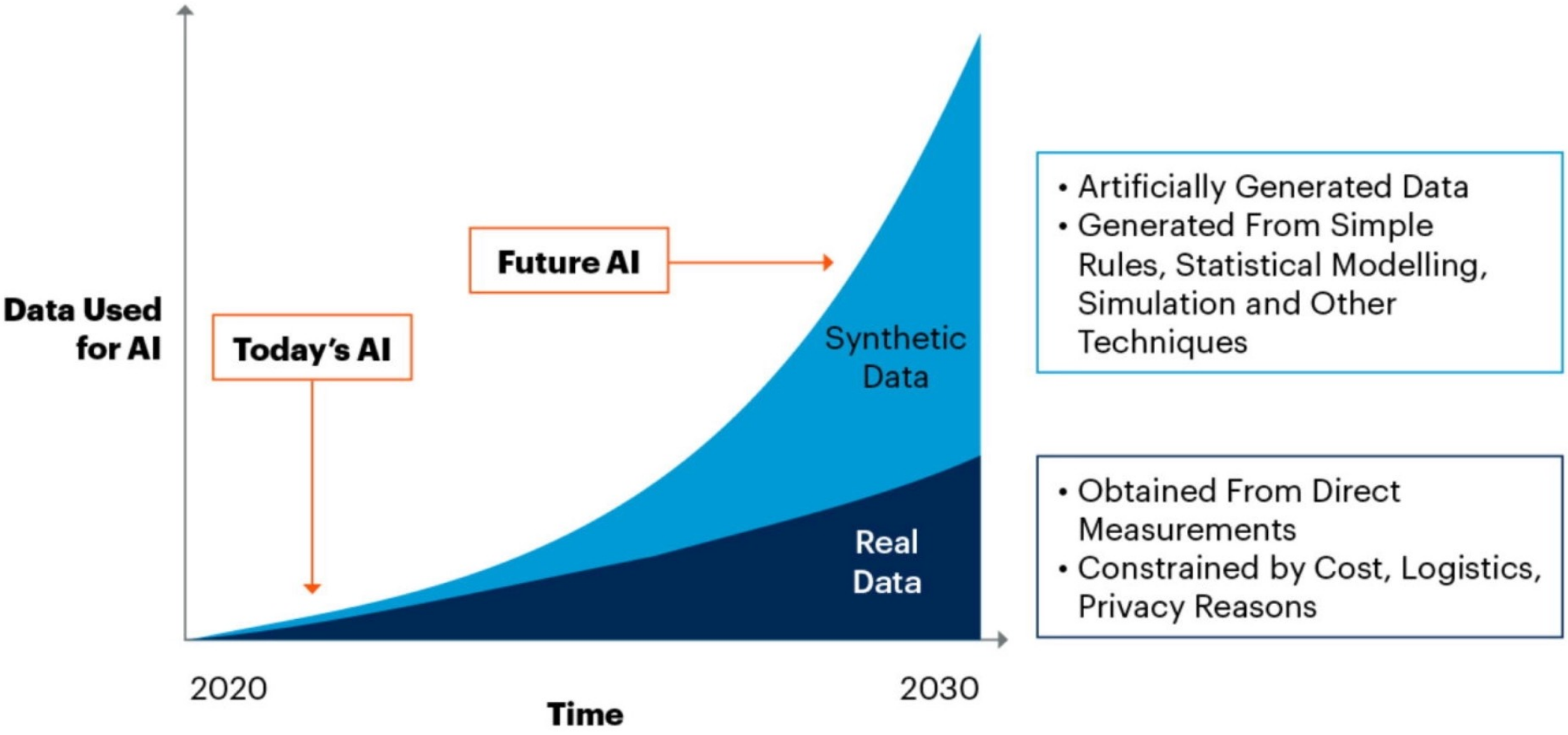
Face simulation

Embodied Perception

Autonomous Driving

- In addition to privacy, synthetic data can address the issue of training data scarcity for certain applications

By 2030, Synthetic Data Will Completely Overshadow Real Data in AI Models



Source: Gartner
750175_C

Gartner

Outline:

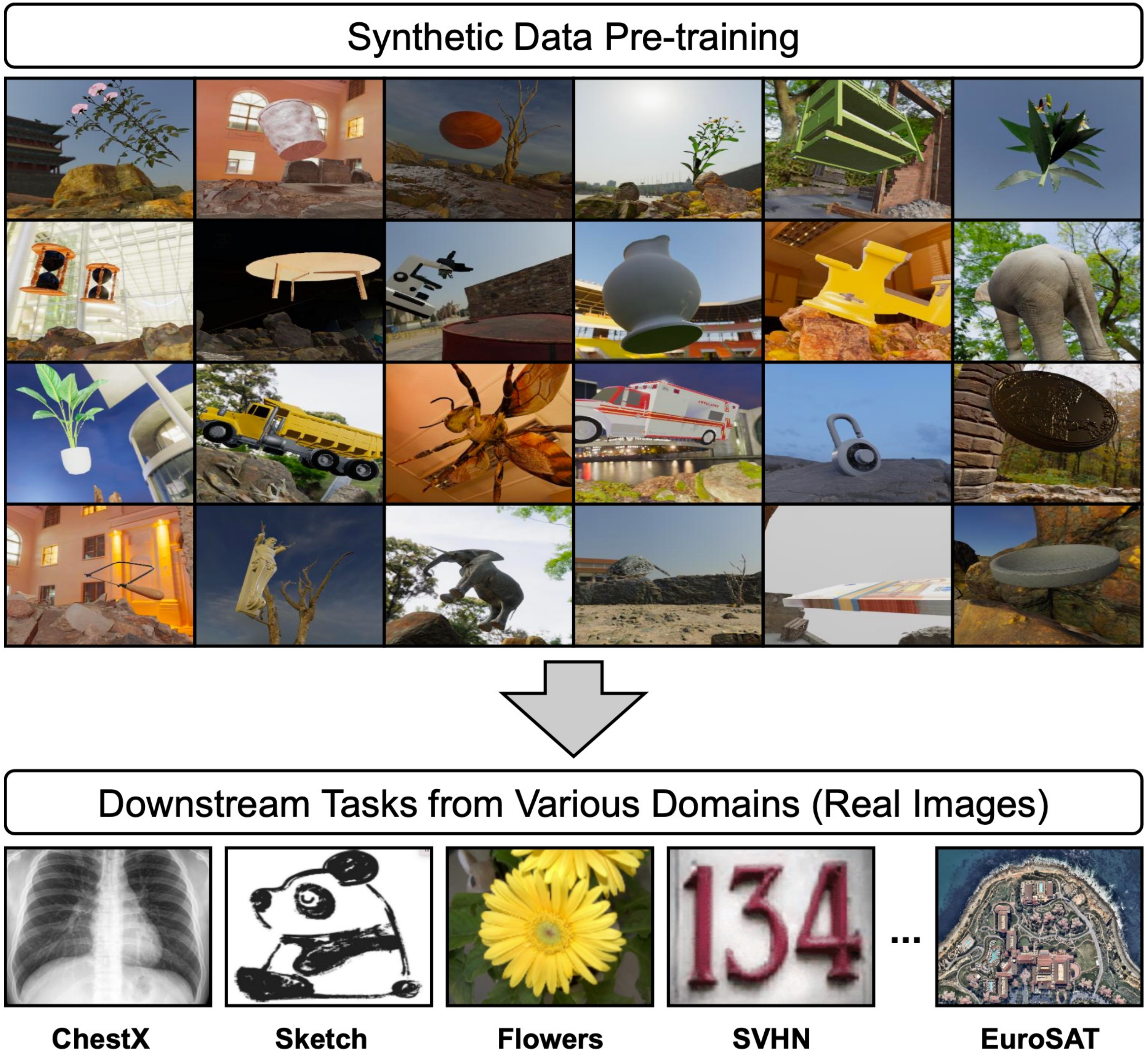
- Pre-training based on Synthetic Data for Vision Tasks
- Synthetic Data for Specialization of Large Language Models
- Application: AI Commentary @ Wimbledon and US Open 2023

Previous Work:

Sim2Real domain adaptation
(same label set)

Our Setting:

New problem: synthetic Data Pretraining and Transfer to Diverse Downstream Tasks
(disjoint label set)



Observation: Different simulation parameters have different effects on different downstream tasks

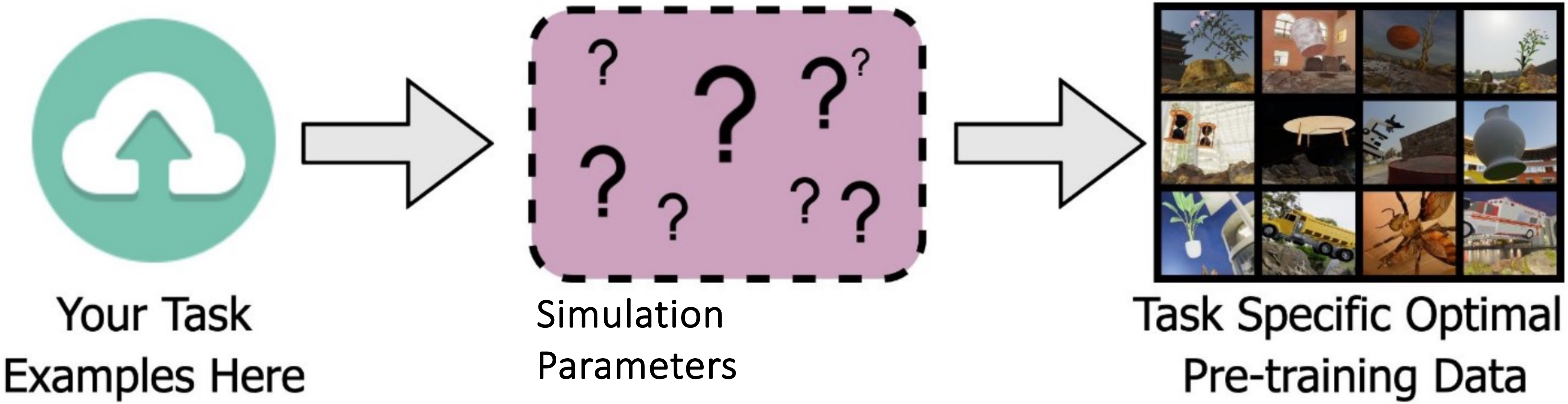
Resnet-50, linear probing

Pretraining Data Variations	Downstream Accuracy			
	EuroSAT	SVHN	Sketch	DTD
Pose	87.01	28.49	37.89	37.39
+Lighting	88.57	32.36	**38.81**	40.32
+Blur	↑ **90.20**	35.58	35.53	37.66 ↓
+Materials	84.54 ↓	**44.84**	30.81	↑ **38.51**
+Background	80.44	29.93	14.60	32.39

Suggests best parameter settings for synthetic data are task-specific

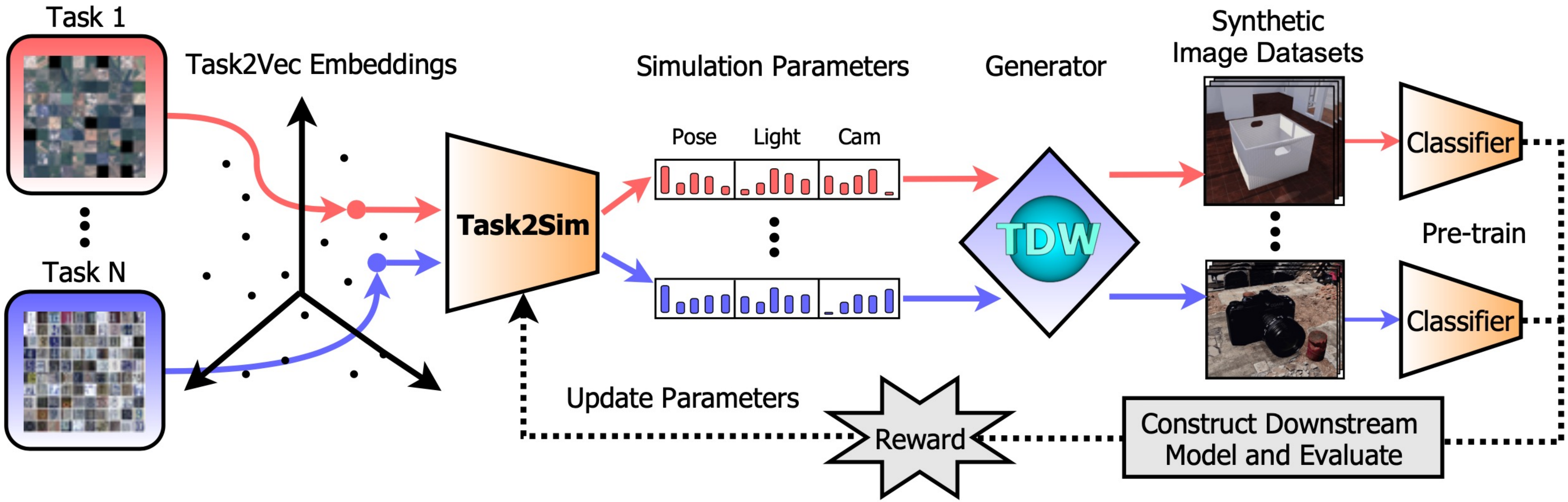
Task2Sim: Towards Effective Pre-training and Transfer from Synthetic Data (CVPR 2022)

S. Mishra, R. Panda, C. Phoo, R. Chen, L. Karlinsly, K. Saenko, V. Saligrama, and R. Feris



Task2Sim: Towards Effective Pre-training and Transfer from Synthetic Data (CVPR 2022)

S. Mishra, R. Panda, C. Phoo, R. Chen, L. Karlinsly, K. Saenko, V. Saligrama, and R. Feris



Task2Sim: Simulation Parameters

Data parameterization : 8 discrete-valued binary scene parameters controlling variation

Cam distance

Materials

Background

Focus Blur

Light color

Obj Rotation

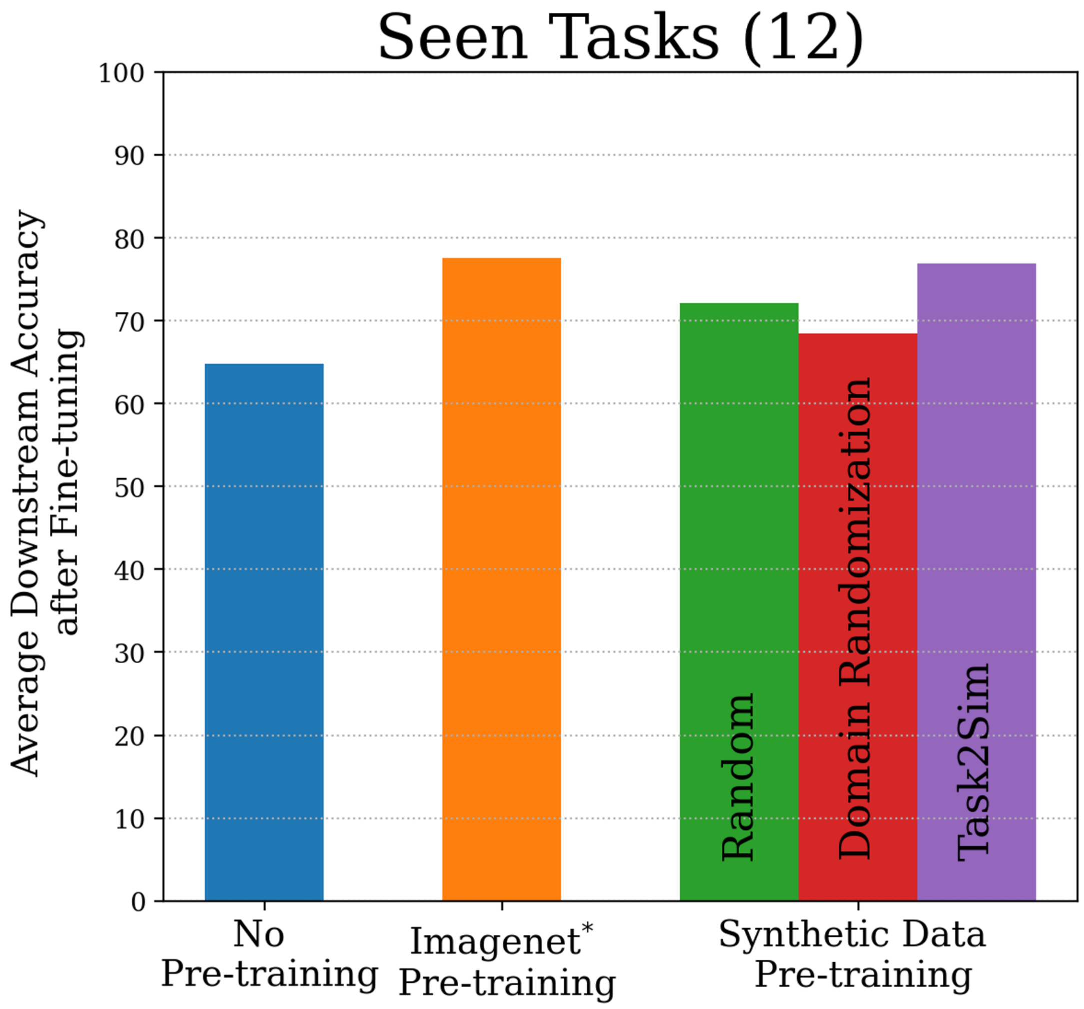
Light Direction

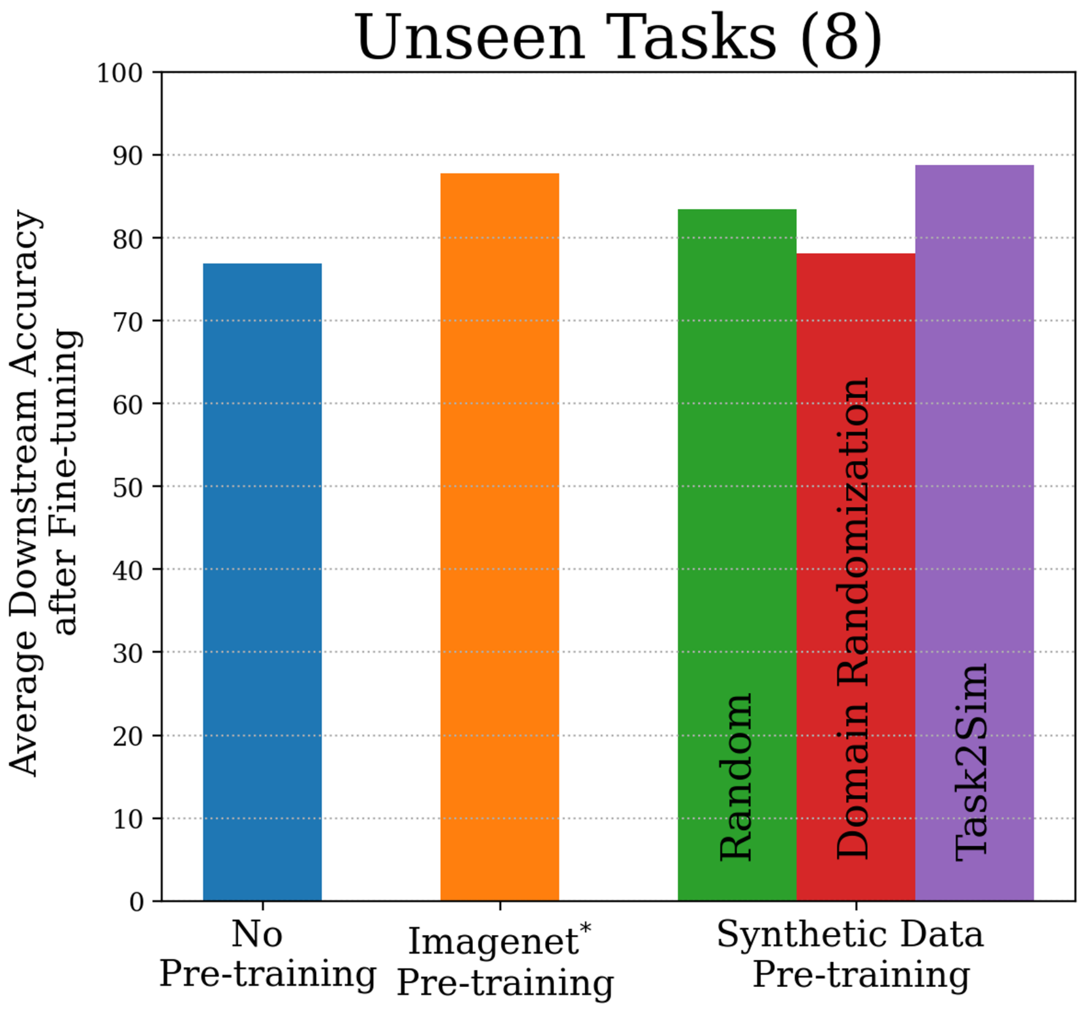
Light Intensity

Experiments:
20 downstream tasks from various domains

Category	Dataset	Train Size	Test Size	Classes
Natural	CropDisease [39]	43456	10849	38
	Flowers [42]	1020	6149	102
	DeepWeeds [44]	12252	5257	9
	CUB [65]	5994	5794	200
Satellite	EuroSAT [18]	18900	8100	10
	Resisc45 [4]	22005	9495	45
	AID [75]	6993	3007	30
	CactusAerial [34]	17500	4000	2
Symbolic	Omniglot [30]	9226	3954	1623
	SVHN [40]	73257	26032	10
	USPS [21]	7291	2007	10
Medical	ISIC [7]	7007	3008	7
	ChestX [67]	18090	7758	7
	ChestXPneumonia [25]	5216	624	2
Illustrative	Kaokore [60]	6568	821	8
	Sketch [66]	35000	15889	1000
	PACS-C [32]	2107	237	7
	PACS-S [32]	3531	398	7
Texture	DTD [6]	3760	1880	47
	FMD [81]	1400	600	10

Task2Sim Downstream Performance





Task2Sim matches ImageNet pre-training and generalizes to tasks not encountered in training

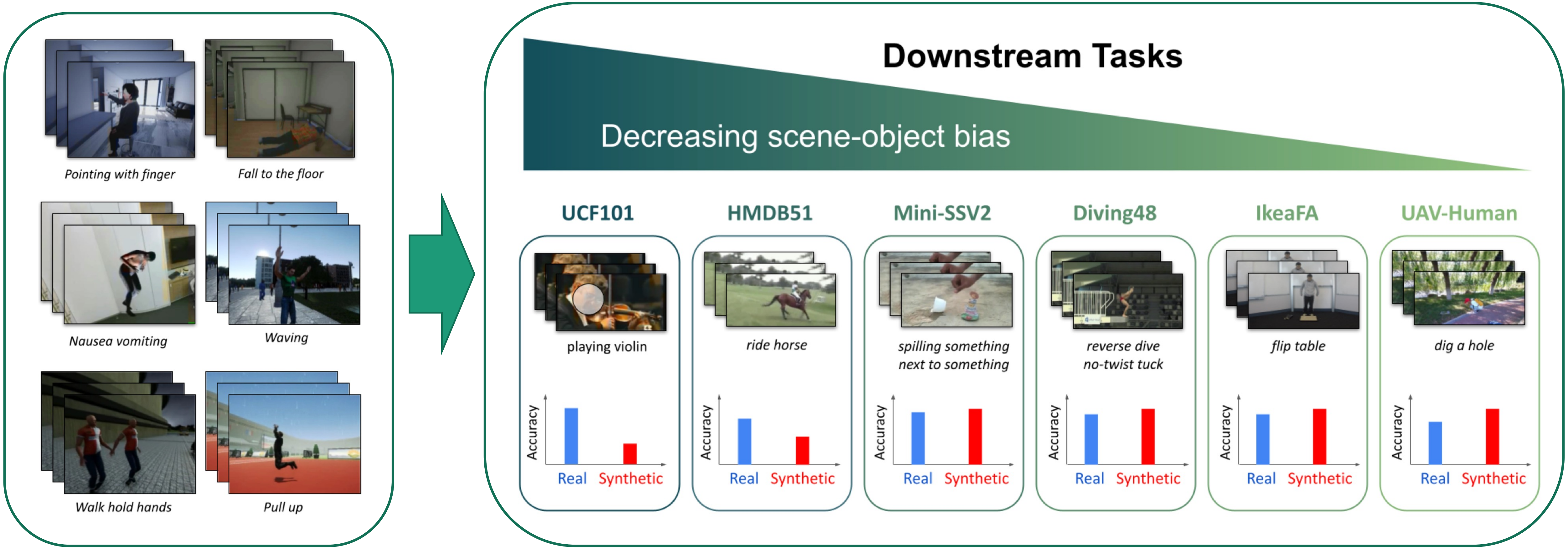
*Indicates subset of Imagenet with same number of total images and total object classes as in synthetic data

How Transferable are Video Representations Based on Synthetic Data?

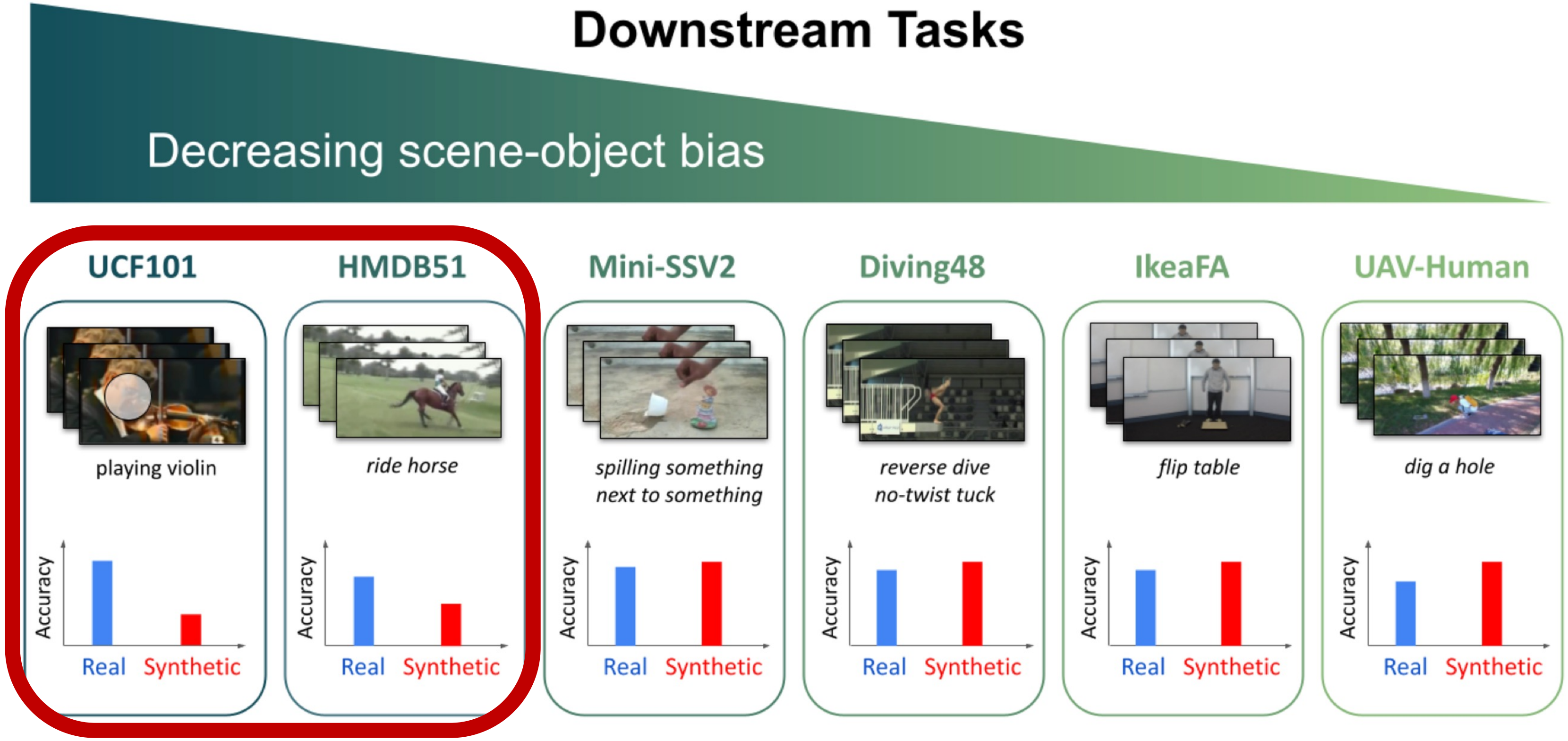
NeurIPS 2022 Dataset Track

Yo-whan Kim, Samarth Mishra, SouYoung Jin, Rameswar Panda, Hilde Kuehne, Leonid Karlinsky,
Venkatesh Saligrama, Kate Saenko, Aude Oliva, Rogerio Feris

Synthetic Pretraining

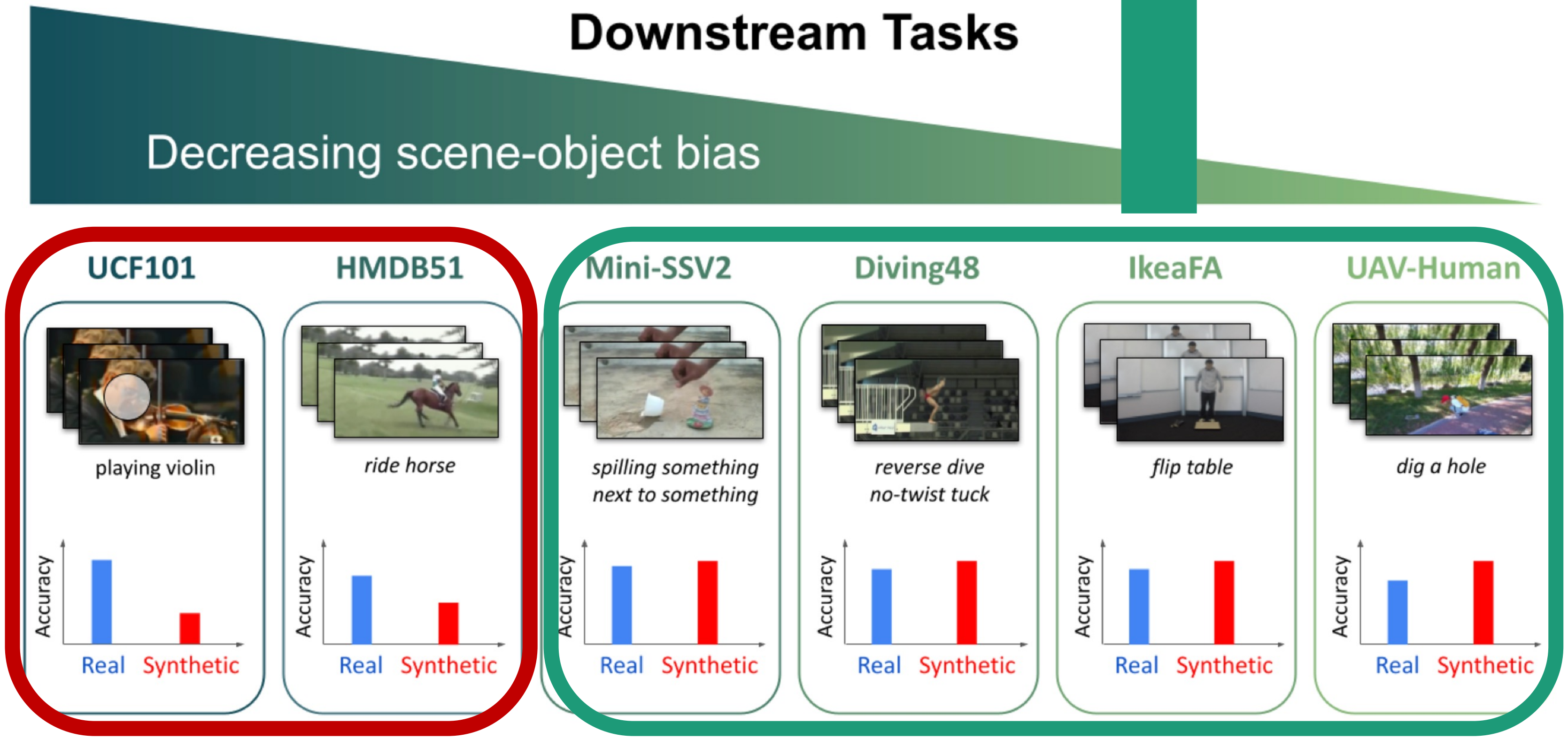


Real pre-training outperforms synthetic pre-training on datasets with high bias.



* Sim2real gap is largely due to contextual features (instead of temporal dynamics of the actions)

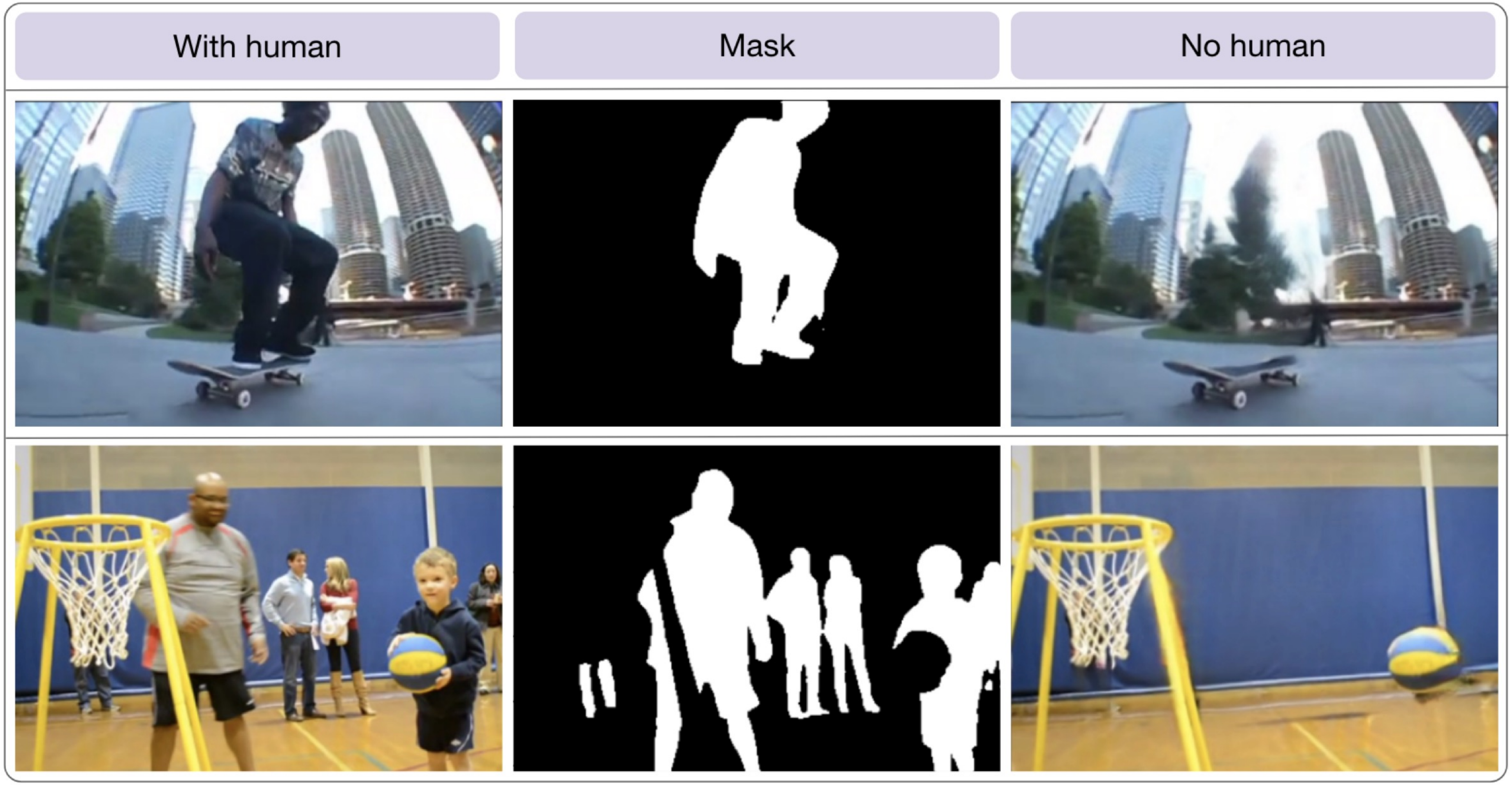
Synthetic pre-training achieves similar or better accuracy than real pre-training on datasets with low bias

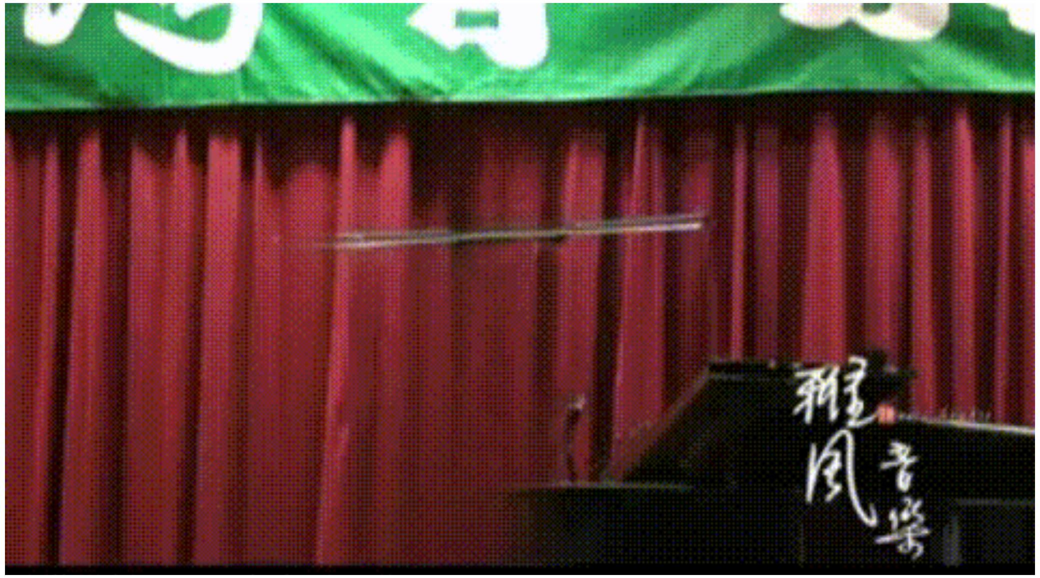
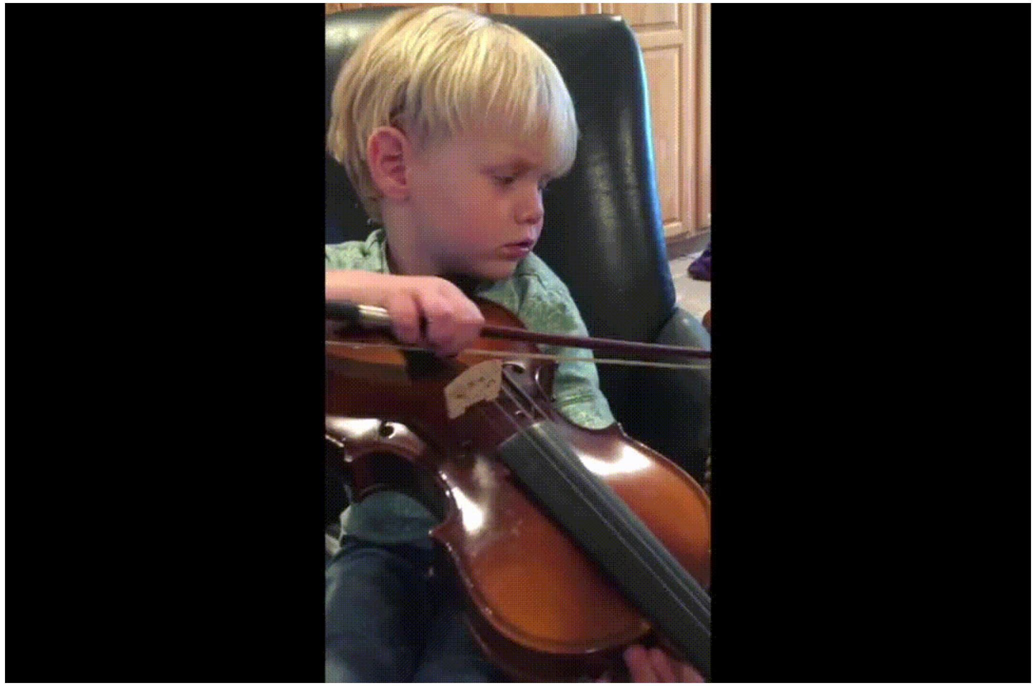
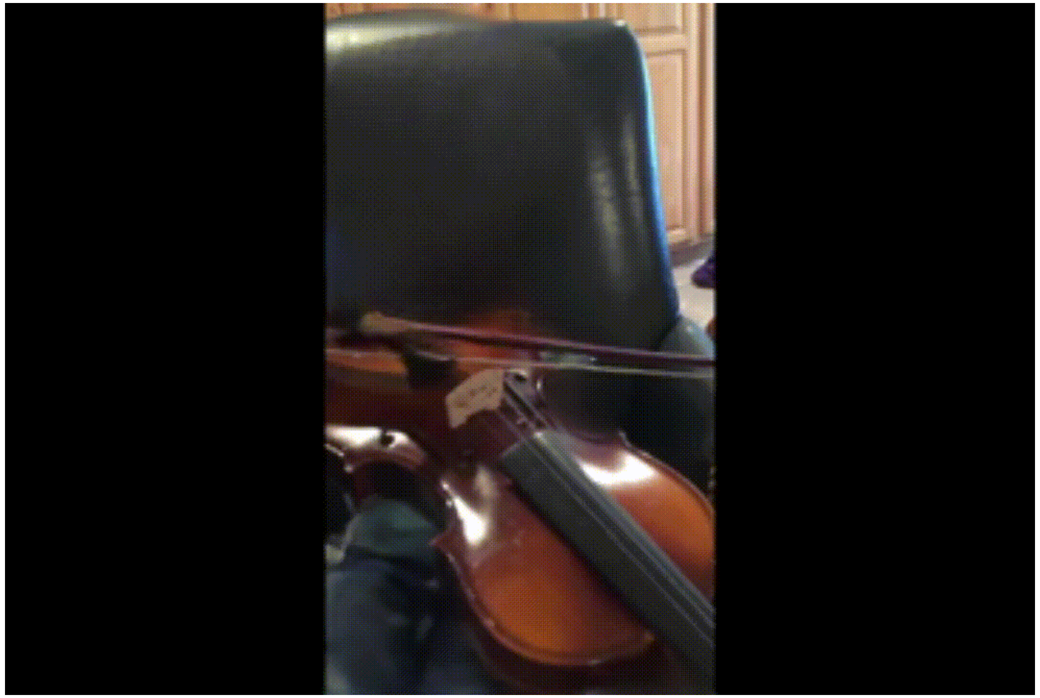


Real pre-training outperforms synthetic pre-training on datasets with high bias

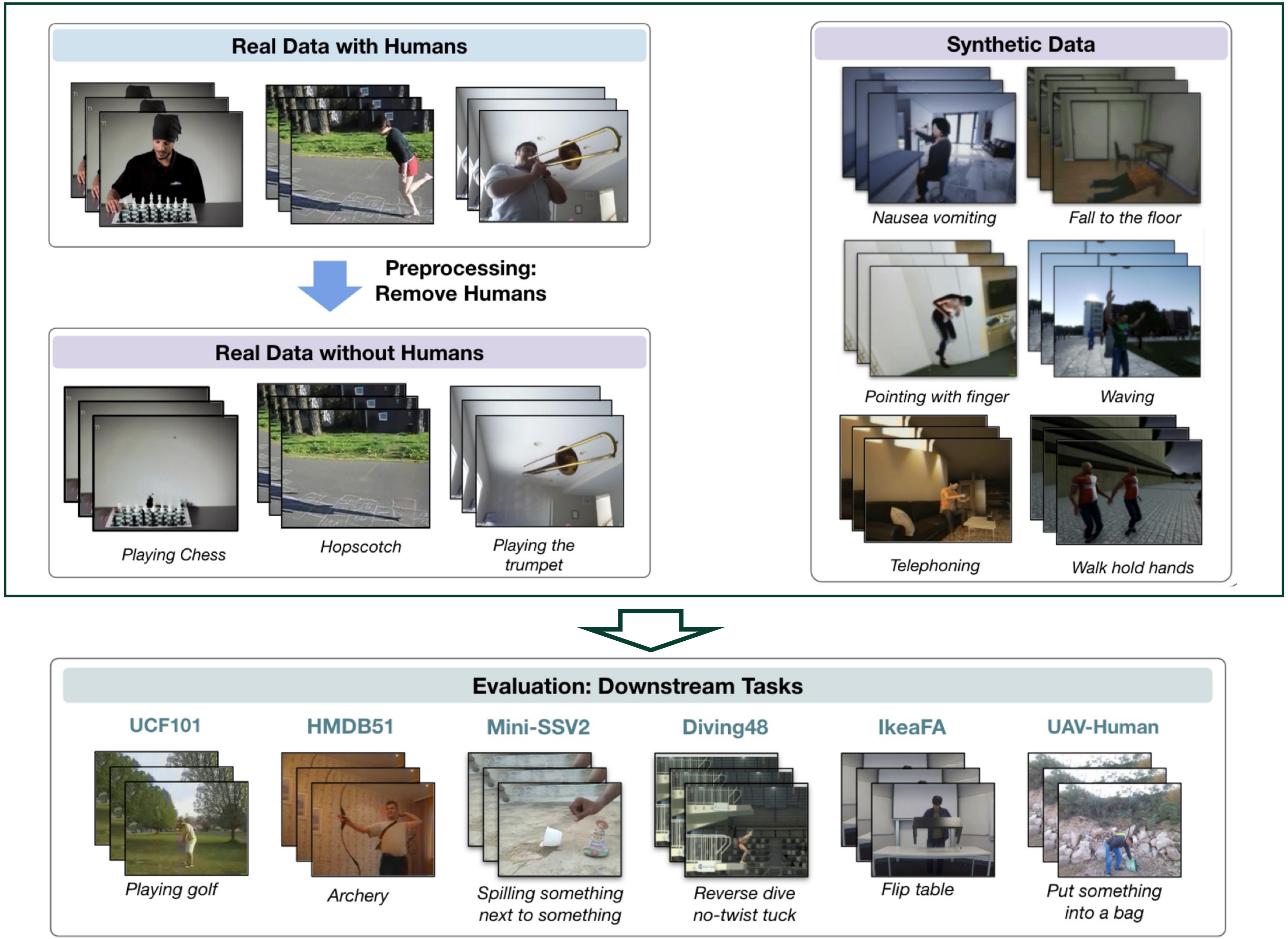
How can we better model contextual features while preserving privacy?

Learning Human Action Recognition Representations Without Humans [Howard Zhong et al, 2023]

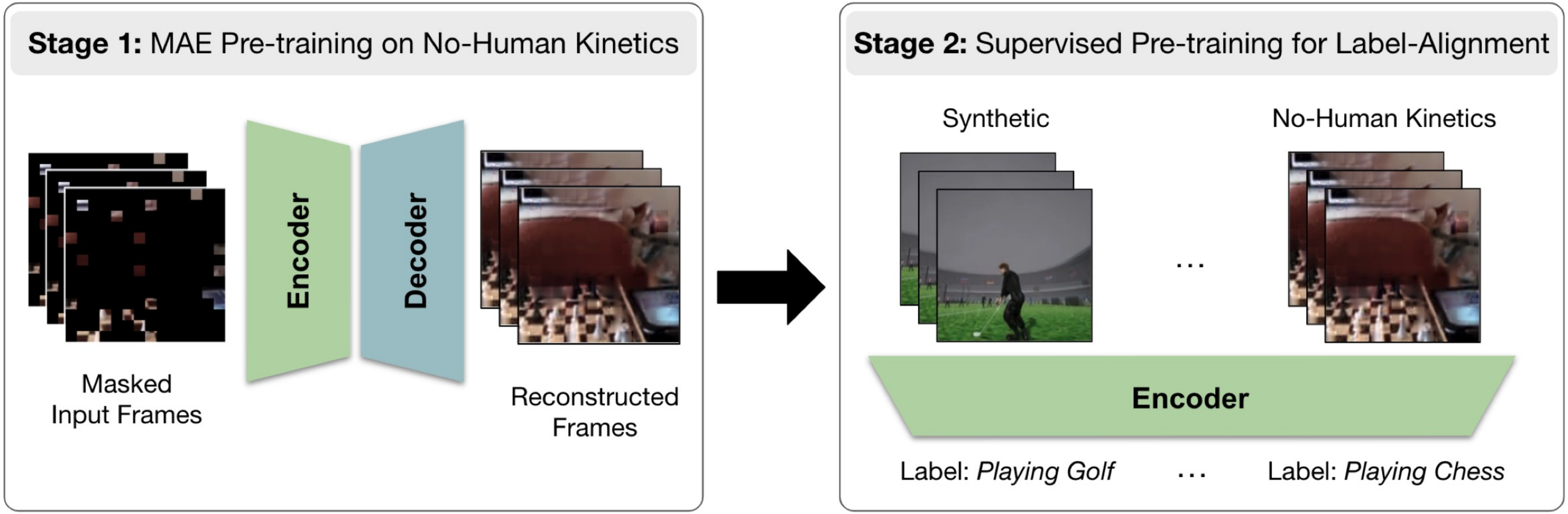


Pre-training: Real Data (no Humans) + Synthetic Data (Virtual Humans)



Privacy-Preserving MAE-Align Pre-training



Experimental Results

Our approach outperforms previous baselines by up to 5% and closes the performance gap between human and no-human action recognition representations

Pre-trained Model	Privacy Preserving	Stage 1: MAE	Stage 2: Alignment	Downstream Task Accuracy												Average	
				UCF101		HMDB51		Mini-SSV2		Diving48		IkeaFA		UAV-Human			
				FT	LP	FT	LP	FT	LP	FT	LP	FT	LP	FT	LP	FT	LP
MAE-Align w. real humans	✖	Kinetics	Kinetics	**93.3**	**91.4**	**73.4**	**69.5**	**68.8**	**37.4**	**66.3**	**19.9**	**72.0**	**58.3**	**34.9**	**13.9**	**68.1**	**48.4**
TimeSformer Kinetics*†	✖	x	Kinetics	92.1	89.4	59.5	55.4	48.9	21.5	46.4	17.0	61.9	47.7	23.3	8.4	55.3	39.9
TimeSformer Synthetic*†	✖	x	Synthetic	89.0	82.1	54.4	49.2	51.1	21.2	44.9	19.2	63.6	45.5	25.0	13.8	54.7	38.5
Scratch	✔	x	x	30.1	-	14.8	-	16.0	-	9.3	-	19.5	-	0.7	-	15.1	-
TSN (RN50 backbone)†	✔	x	Synthetic	83.4	28.0	54.4	20.9	49.7	12.8	63.5	10.9	42.7	36.0	35.6	5.7	54.9	19.1
I3D (RN50 backbone)†	✔	x	Synthetic	82.1	27.6	55.7	22.6	50.7	12.3	55.3	10.1	42.7	33.2	35.1	5.8	53.6	18.6
R(2+1)D (RN50 backbone)†	✔	x	Synthetic	80.0	26.4	53.3	22.2	52.0	13.3	57.3	10.0	41.5	35.7	31.8	5.5	52.6	18.9
MAE-Align w. Synthetic	✔	Synthetic	Synthetic	88.7	76.6	69.7	59.7	64.3	26.0	61.1	16.7	67.3	**57.7**	36.1	**20.6**	64.5	42.9
Ours : Privacy-Preserving MAE-Align (PPMA)	✔	NH Kinetics	NH Kinetics + Synthetic	**92.5**	**88.4**	**71.2**	**64.9**	**67.8**	**34.9**	**64.0**	**21.9**	**67.9**	**57.7**	**38.5**	19.3	**67.0**	**47.9**

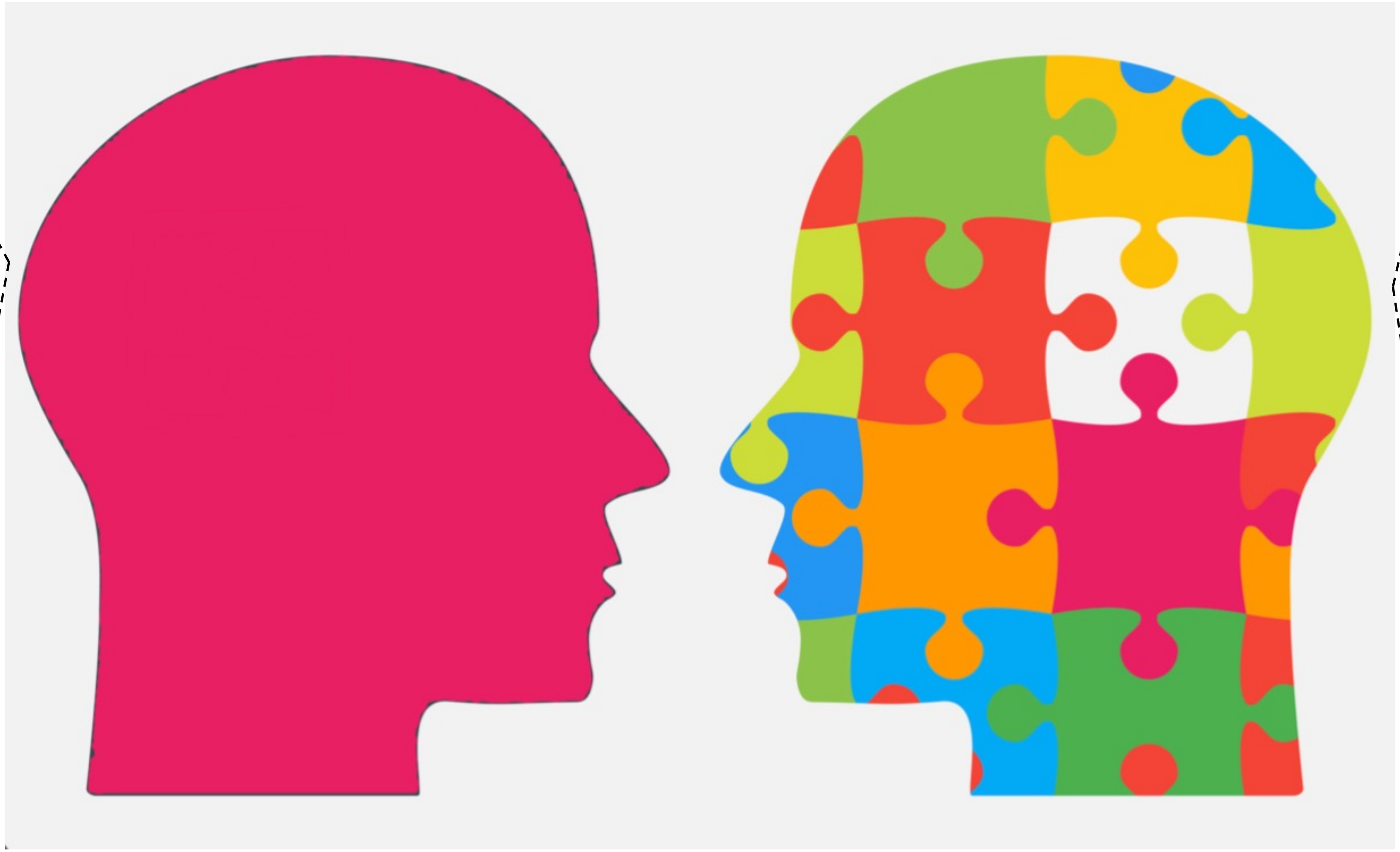
Outline:

- Pre-training based on Synthetic Data for Vision Tasks
- Synthetic Data for Specialization of Large Language Models
- Application: AI Commentary @ Wimbledon and US Open 2023

Generalist

- Does well in many domains
- Requires more capacity (bigger models)

Specialist | Generalist



- <u>Best</u> in a specific domain
- Requires less capacity (smaller models)

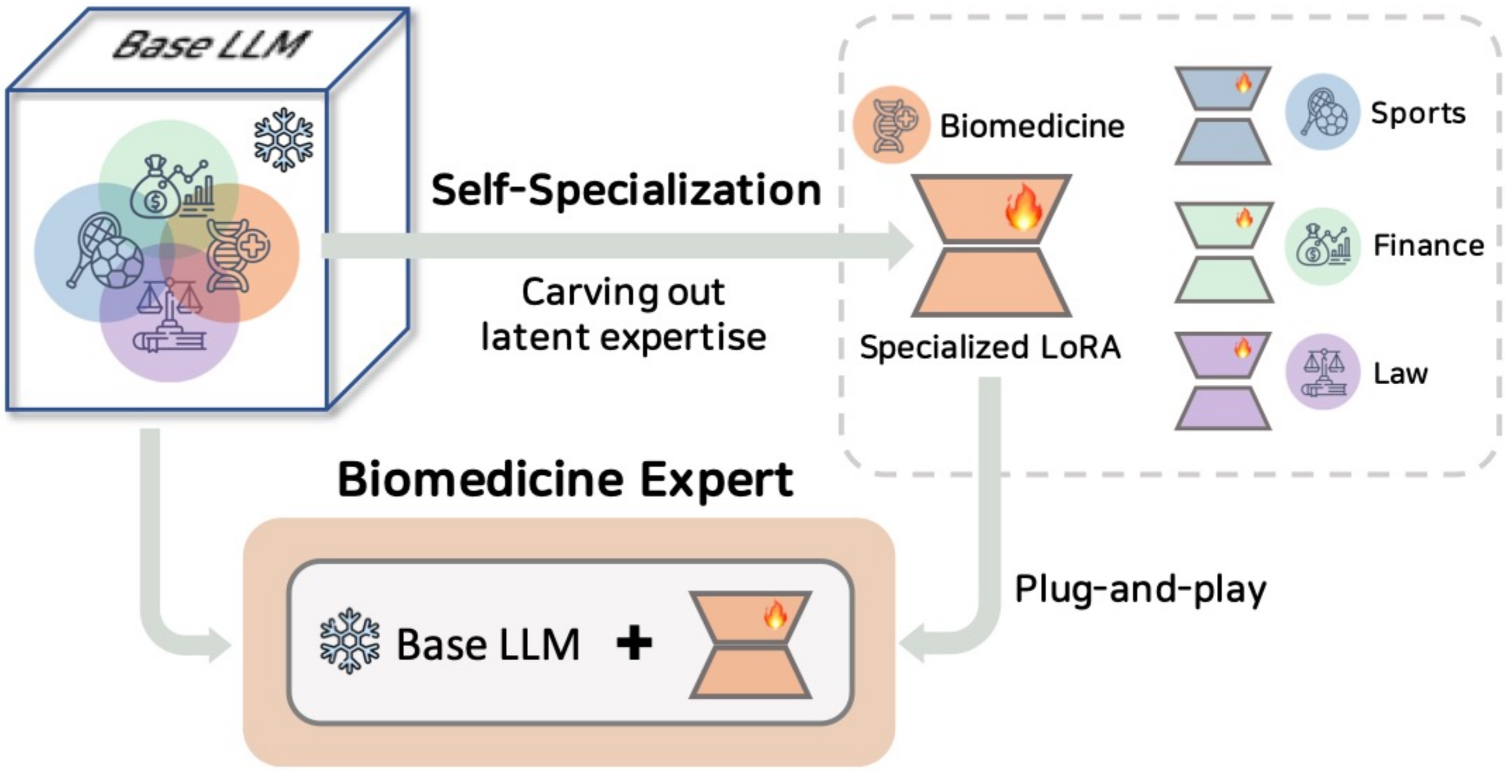
- Does well in many domains
- Requires more capacity (bigger models)

Can we adapt a general large language model to perform better in a specific domain?

- Naive Approach: Instruction fine-tuning, but it requires time-consuming and difficult to scale manual annotation

SELF-SPECIALIZATION: UNCOVERING LATENT EXPERTISE WITHIN LARGE LANGUAGE MODELS

Junmo Kang[*1] **Hongyin Luo**[2] **Yada Zhu**[3] **James Glass**[2] **David Cox**[3]
Alan Ritter[1] **Rogerio Feris**[3] **Leonid Karlinsky**[3]



Real-Data:

Domain-specific seed instructions (small set – e.g., 80)

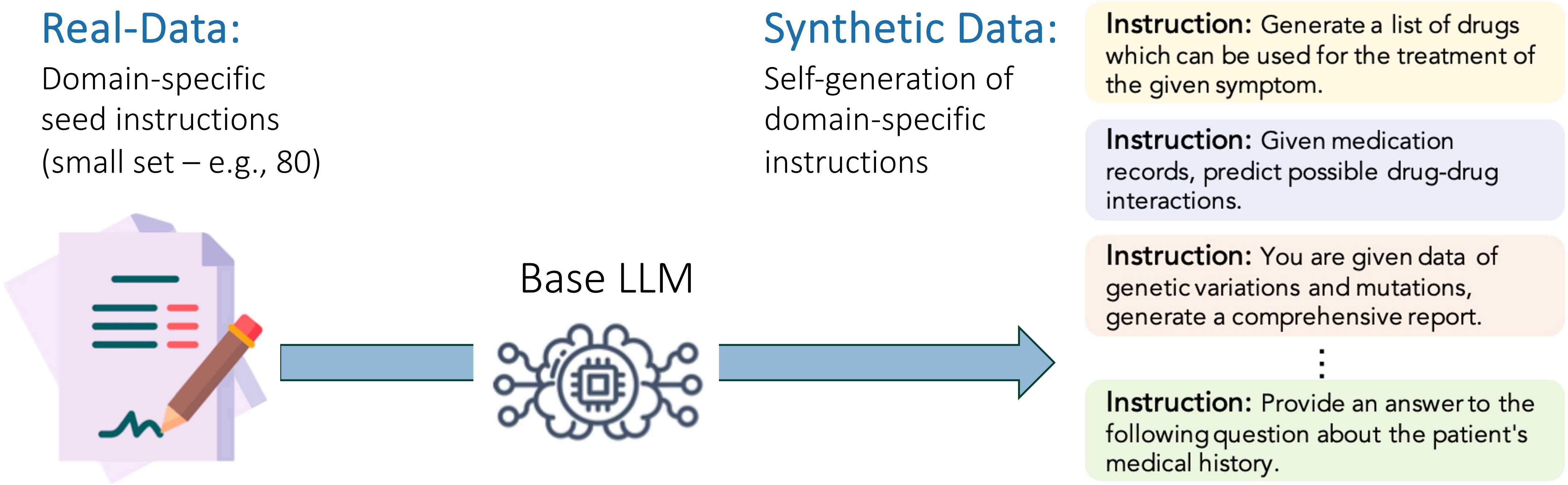
Instruction:

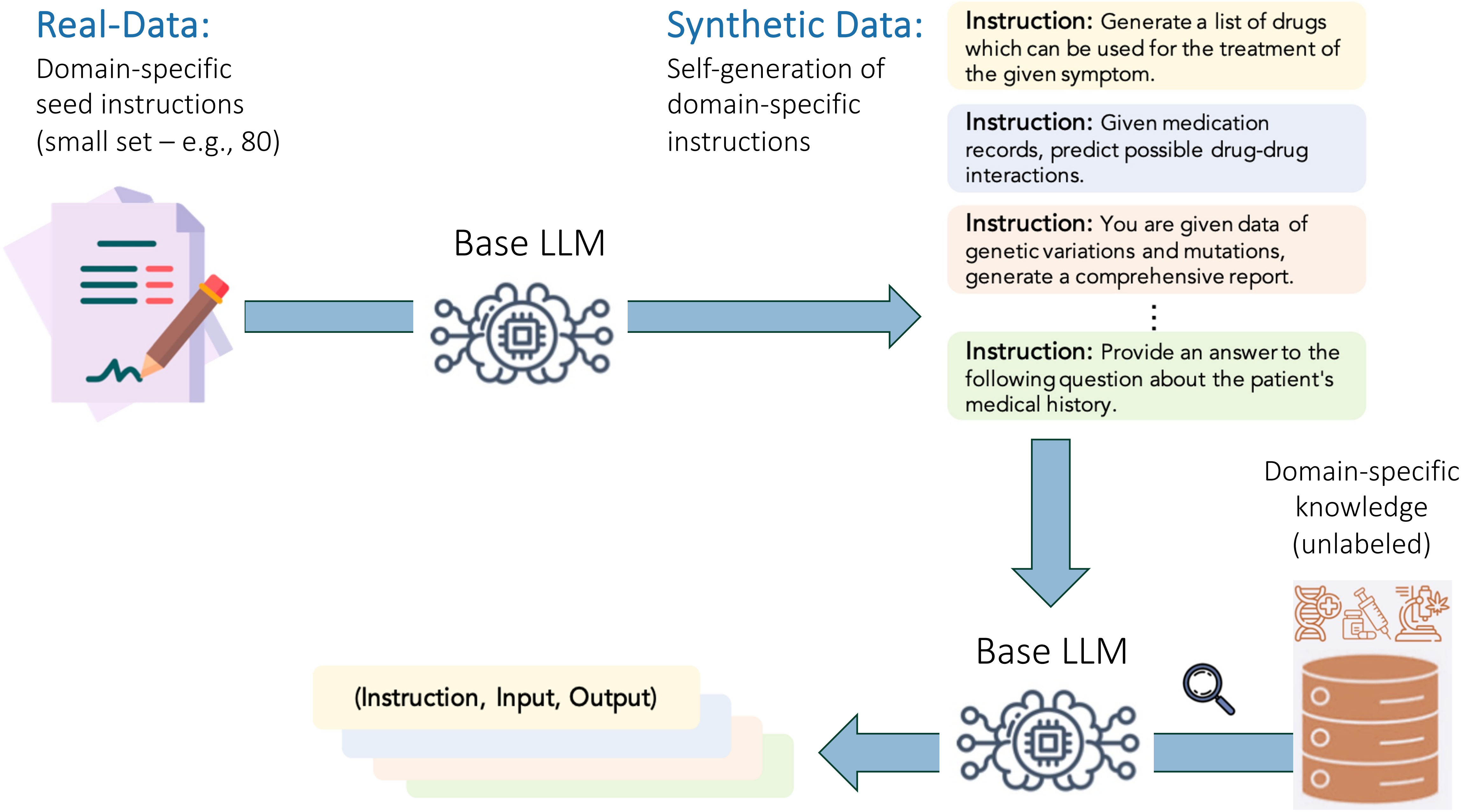
In this task, you are given a short article and question. Read the short article and answer the question.

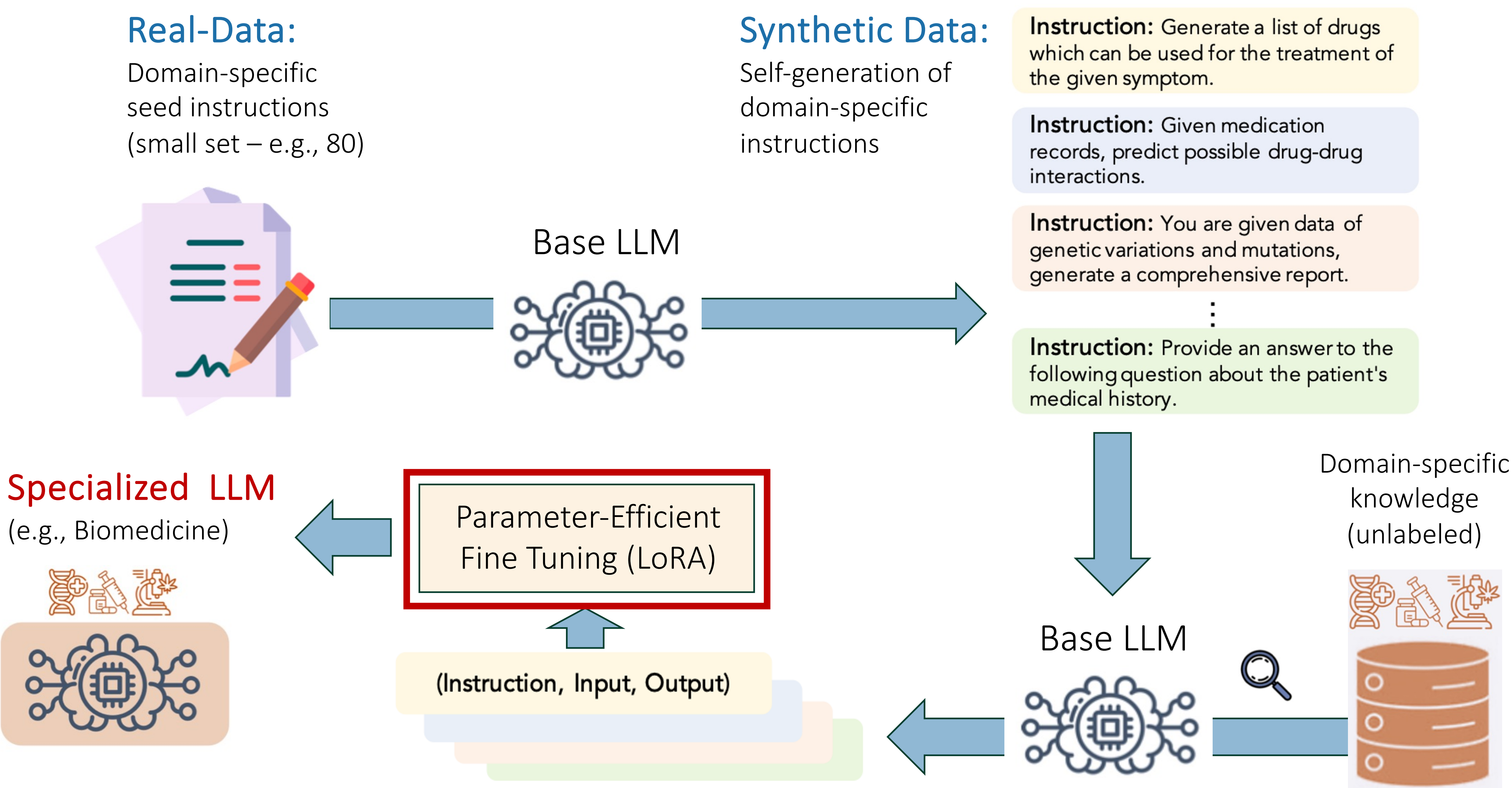
Input:

Short article: The extract from the opium poppy was tested on breast cancer cells and was found to inhibit the migration and invasion of breast cancer cells.

Question: Was the extract more potent in its inhibitory effect on the migration of breast cancer cells than its effect on the invasion of breast cancer cells?


Real-Data:
Domain-specific seed instructions (small set – e.g., 80)
Base LLM
Synthetic Data:
Self-generation of domain-specific instructions
Instruction: Generate a list of drugs which can be used for the treatment of the given symptom.
Instruction: Given medication records, predict possible drug-drug interactions.
Instruction: You are given data of genetic variations and mutations, generate a comprehensive report.
⋮
Instruction: Provide an answer to the following question about the patient's medical history.


Real-Data:
Domain-specific seed instructions (small set – e.g., 80)
Base LLM
Synthetic Data:
Self-generation of domain-specific instructions
Instruction: Generate a list of drugs which can be used for the treatment of the given symptom.
Instruction: Given medication records, predict possible drug-drug interactions.
Instruction: You are given data of genetic variations and mutations, generate a comprehensive report.
⋮
Instruction: Provide an answer to the following question about the patient's medical history.
Domain-specific knowledge (unlabeled)
Base LLM
(Instruction, Input, Output)


Real-Data:
Domain-specific seed instructions (small set – e.g., 80)
Base LLM
Synthetic Data:
Self-generation of domain-specific instructions
Instruction: Generate a list of drugs which can be used for the treatment of the given symptom.
Instruction: Given medication records, predict possible drug-drug interactions.
Instruction: You are given data of genetic variations and mutations, generate a comprehensive report.
Instruction: Provide an answer to the following question about the patient's medical history.
Domain-specific knowledge (unlabeled)
Base LLM
(Instruction, Input, Output)
Parameter-Efficient Fine Tuning (LoRA)
Specialized LLM
(e.g., Biomedicine)

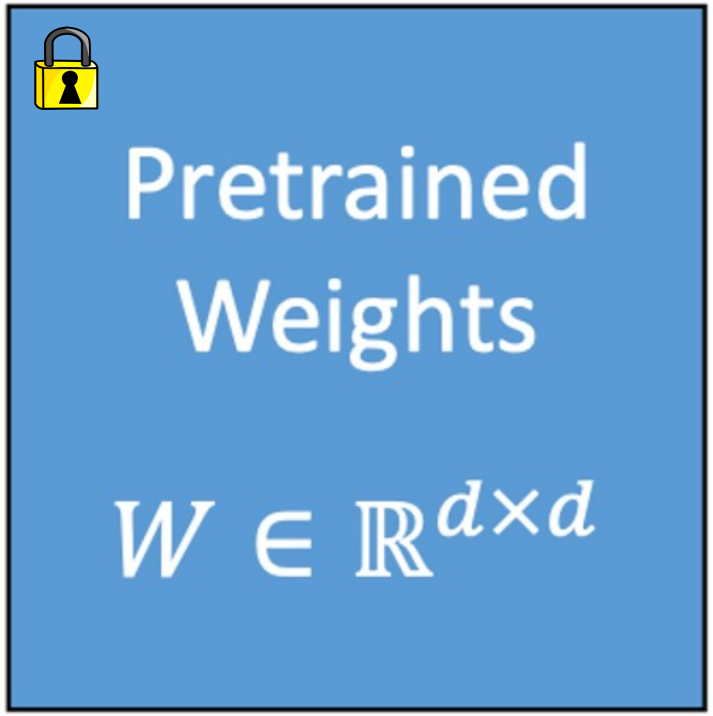
Traditional Fine-tuning

- All pre-trained model weights are optimized
- Not feasible for billion-scale models

LoRA: Low-rank adaptation of large language models

[Hu et al, ICLR 2022]

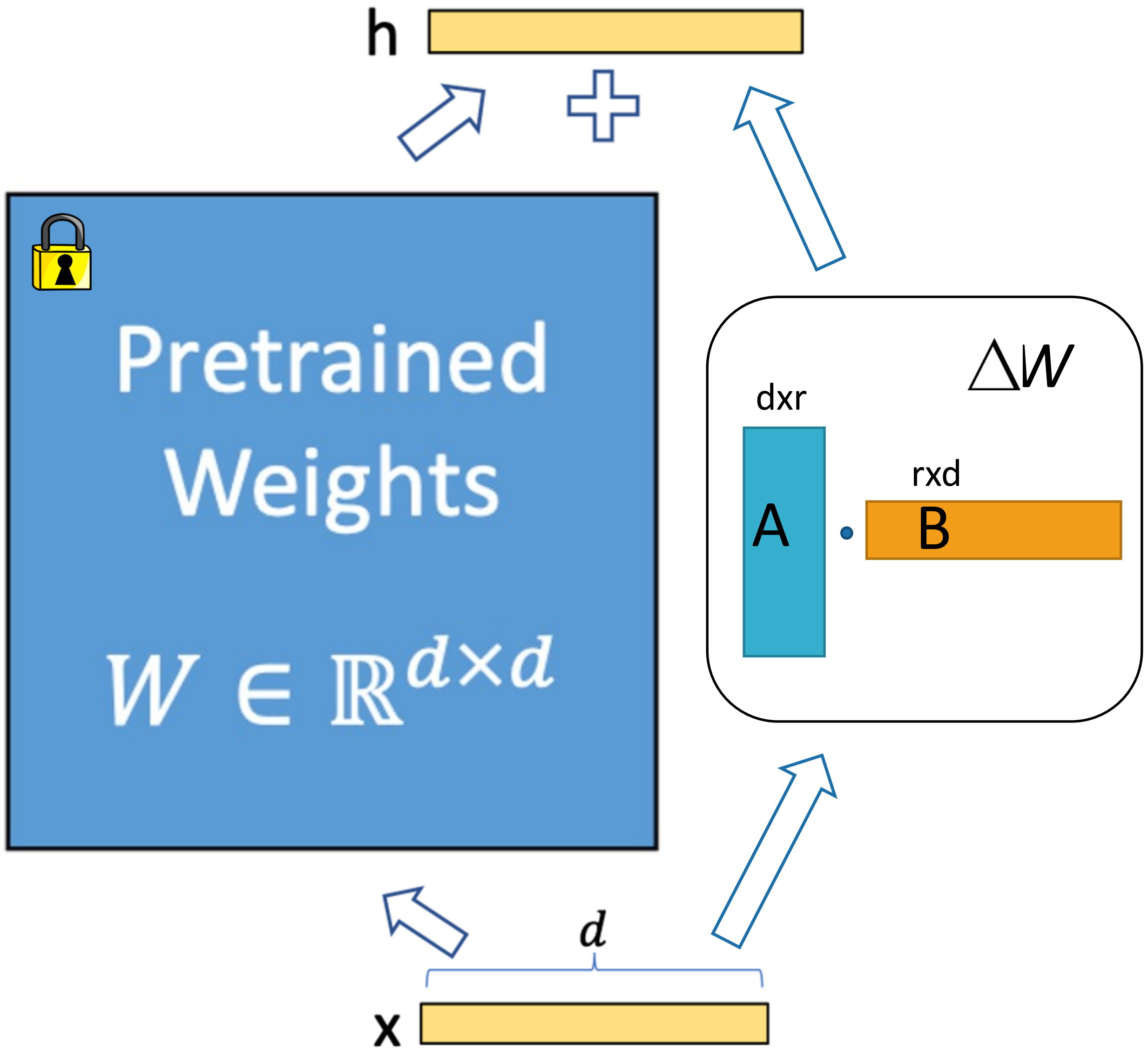
Pre-trained weights are frozen



LoRA: Low-rank adaptation of large language models

[Hu et al, ICLR 2022]

Pre-trained weights are frozen

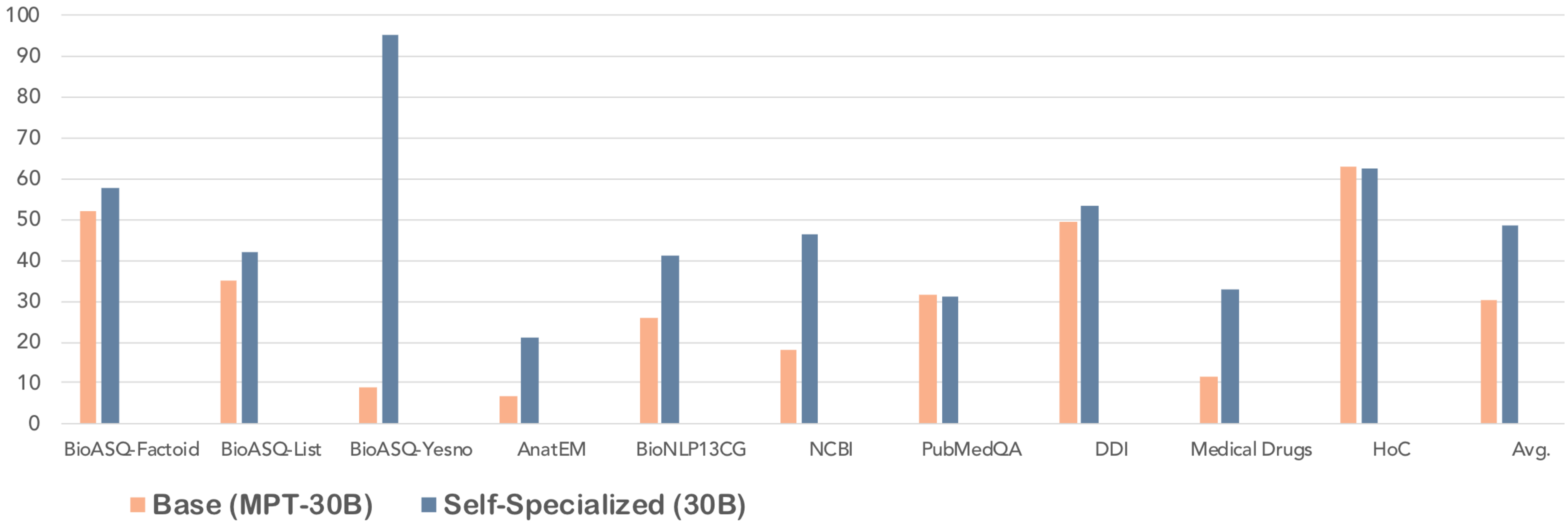


Only optimize A and B

Main Results

- Our self-specialized model (30B) outperforms its base model, MPT-30B by a large margin (up to 18 points)

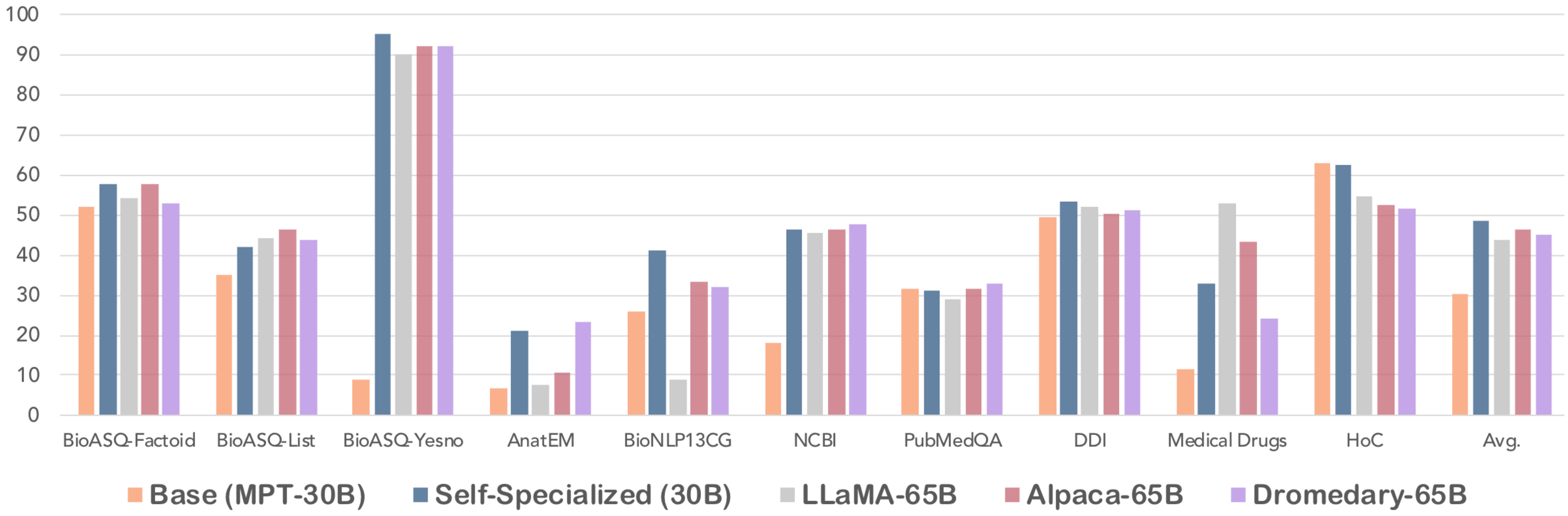
5-shot results in 10 datasets (biomedicine)



Main Results

- Our self-specialized model (30B) even outperforms 65B models (2.2x larger), including Llama-65B

5-shot results in 10 datasets (biomedicine)



Outline:

- Pre-training based on Synthetic Data for Vision Tasks
- Synthetic Data for Specialization of Large Language Models
- Application: AI Commentary @ Wimbledon and US Open 2023

90 players

18 holes

4 days

100s hours of footage

since **1877** **>600** players

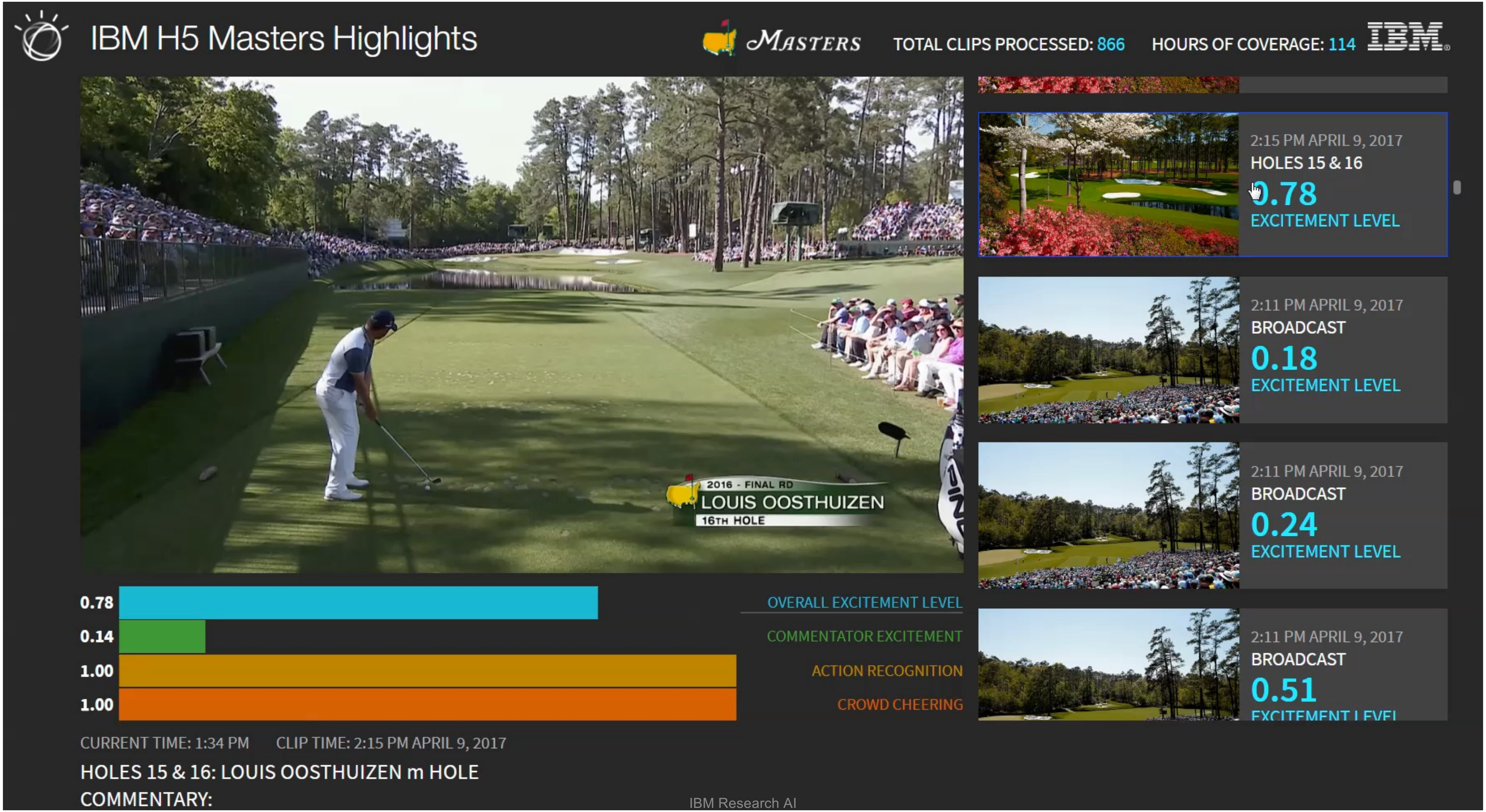
19 courts **13** days **254** matches

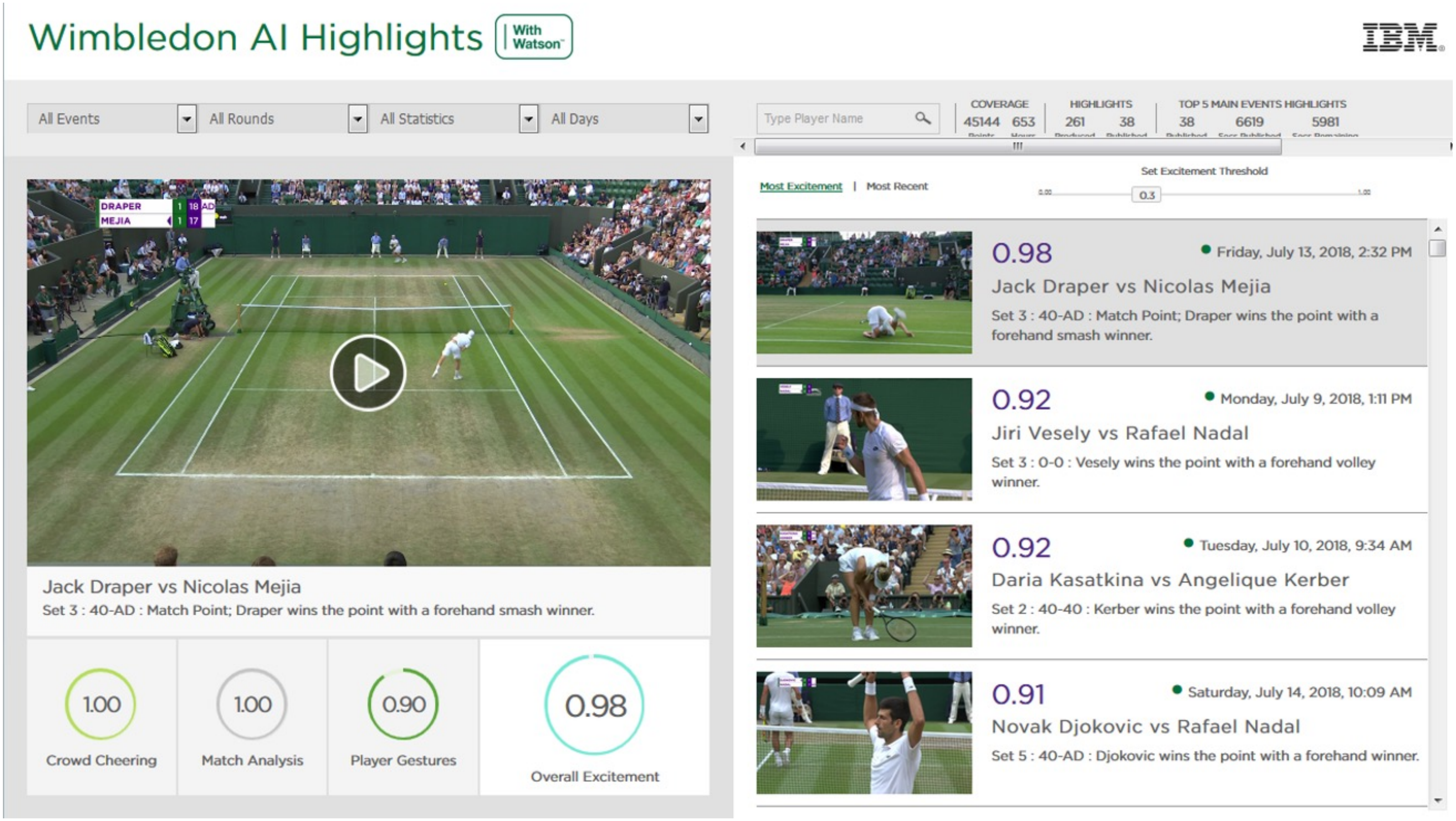
100s hours of footage

14 days **400** players

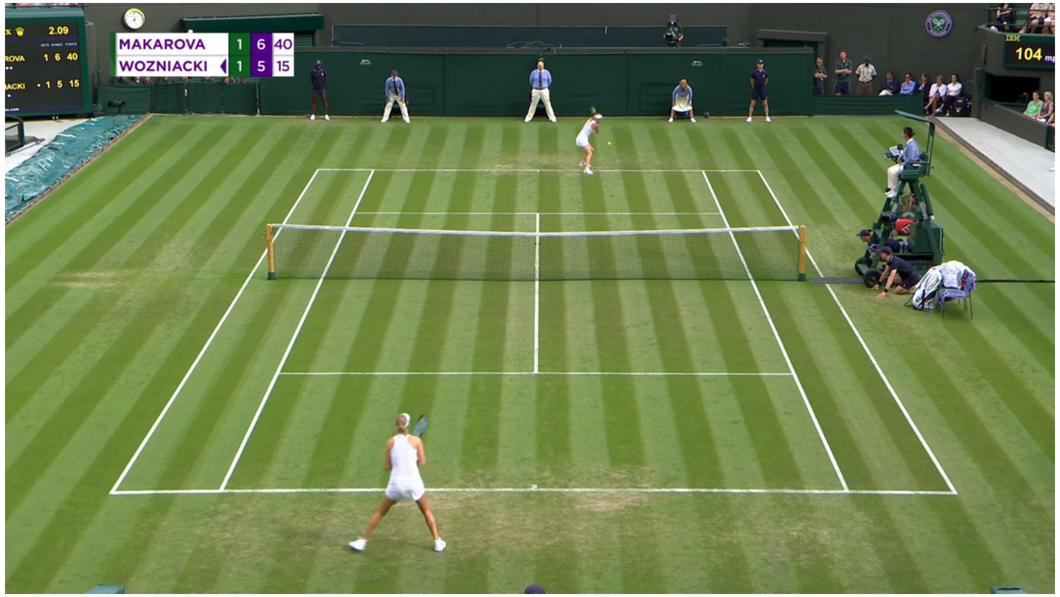
304 hours of coverage

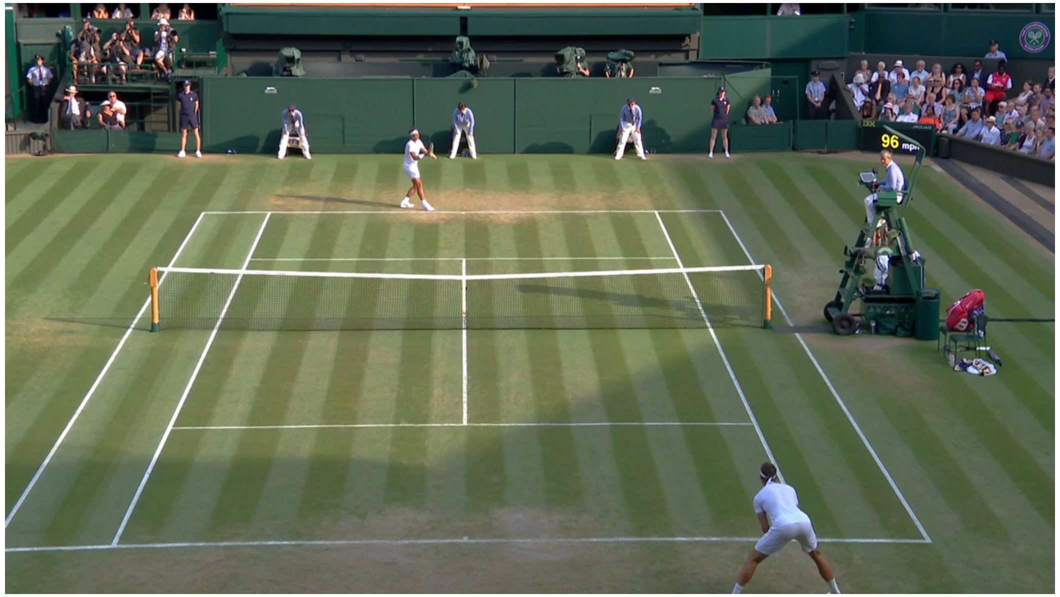
24K clips processed

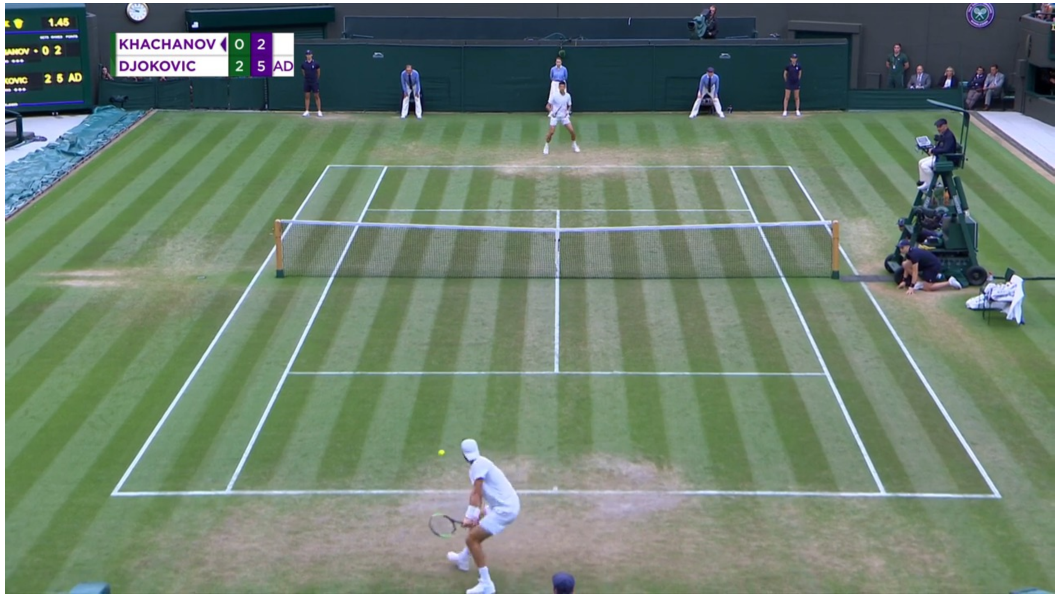

IBM H5 Masters Highlights
Masters
TOTAL CLIPS PROCESSED: 866
HOURS OF COVERAGE: 114
IBM
2016 - FINAL RD
LOUIS OOSTHUIZEN
16TH HOLE
2:15 PM APRIL 9, 2017
HOLES 15 & 16
0.78
EXCITEMENT LEVEL
2:11 PM APRIL 9, 2017
BROADCAST
0.18
EXCITEMENT LEVEL
2:11 PM APRIL 9, 2017
BROADCAST
0.24
EXCITEMENT LEVEL
2:11 PM APRIL 9, 2017
BROADCAST
0.51
0.78
OVERALL EXCITEMENT LEVEL
0.14
COMMENTATOR EXCITEMENT
1.00
ACTION RECOGNITION
1.00
CROWD CHEERING
CURRENT TIME: 1:34 PM
CLIP TIME: 2:15 PM APRIL 9, 2017
HOLES 15 & 16: LOUIS OOSTHUIZEN m HOLE
COMMENTARY:
IBM Research AI

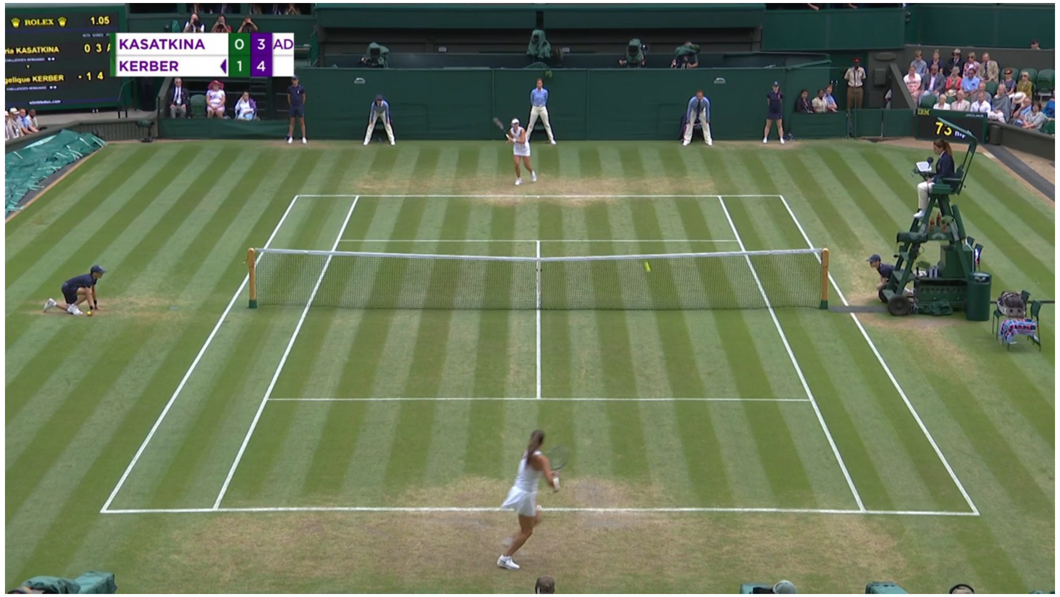

Wimbledon AI Highlights
With Watson
IBM
All Events
All Rounds
All Statistics
All Days
Type Player Name
COVERAGE
45144
653
HIGHLIGHTS
261
38
TOP 5 MAIN EVENTS HIGHLIGHTS
38
6619
5981
Most Excitement | Most Recent
Set Excitement Threshold
0.3
DRAPER 1 18 AD
MEJIA 1 17
Jack Draper vs Nicolas Mejia
Set 3 : 40-AD : Match Point; Draper wins the point with a forehand smash winner.
1.00
Crowd Cheering
1.00
Match Analysis
0.90
Player Gestures
0.98
Overall Excitement
0.98
Friday, July 13, 2018, 2:32 PM
Jack Draper vs Nicolas Mejia
Set 3 : 40-AD : Match Point; Draper wins the point with a forehand smash winner.
0.92
Monday, July 9, 2018, 1:11 PM
Jiri Vesely vs Rafael Nadal
Set 3 : 0-0 : Vesely wins the point with a forehand volley winner.
0.92
Tuesday, July 10, 2018, 9:34 AM
Daria Kasatkina vs Angelique Kerber
Set 2 : 40-40 : Kerber wins the point with a forehand volley winner.
0.91
Saturday, July 14, 2018, 10:09 AM
Novak Djokovic vs Rafael Nadal
Set 5 : 40-AD : Djokovic wins the point with a forehand winner.

Multi-modal curation of sports highlights (Wimbledon)









Impact

- Our system has been used to produce the official highlights of the US Open, Wimbledon, and Masters tournaments
- Watched by millions of fans worldwide (mobile app, website, ...)

The New York Times

Enjoy Those U.S. Open Highlights. A Computer Picked Them for You.

(and many more media outlets)

Our work on AI/ML for auto-curation of sports highlights has been selected for a Technology and Engineering Emmy Award !

NATIONAL ACADEMY OF TELEVISION ARTS & SCIENCES

AI Commentary

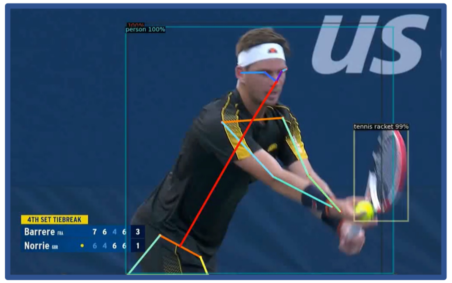
Wimbledon and US Open 2023

Turning multi-sensory data into fluent commentary using Large Language Models

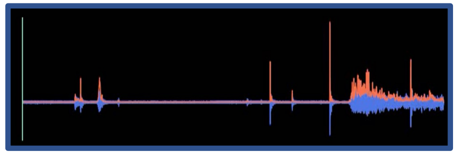
After a wide out serve by Peniston, Laaksonen returns the service with a short backhand. Peniston then approaches the net, hits a forehand, but Laaksonen strikes back with a magnificent backhand. What a beautiful shot! Peniston makes all efforts to save the point but hits the net with a volley unforced error. The game is now tied: 30-30

Billion-scale IBM Language Model

- IBM Sandstone.3B model
- Pretrained on petabytes of data
- Fine-tuned on real and synthetic commentary data



Video



Sound

Match Data

Excitement Scores

Projection into a unified embedding space

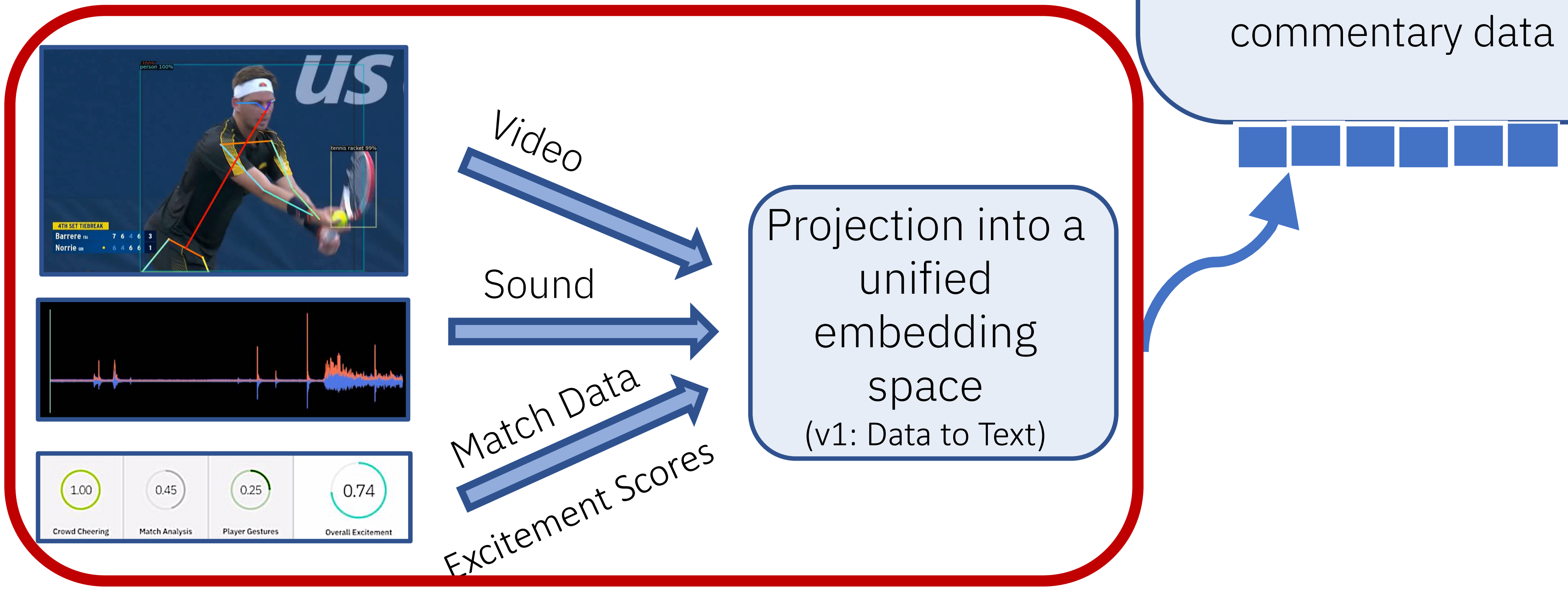
(v1: Data to Text)

Turning multi-sensory data into fluent commentary using Large Language Models

After a wide out serve by Peniston, Laaksonen returns the service with a short backhand. Peniston then approaches the net, hits a forehand, but Laaksonen strikes back with a magnificent backhand. What a beautiful shot! Peniston makes all efforts to save the point but hits the net with a volley unforced error. The game is now tied: 30-30

Billion-scale IBM Language Model

- IBM Sandstone.3B model
- Pretrained on petabytes of data
- Fine-tuned on real and synthetic commentary data



Play-by-play metadata extraction using computer vision

Understanding every detail of the game in real-time

Court and Net Detection

Player Tracking

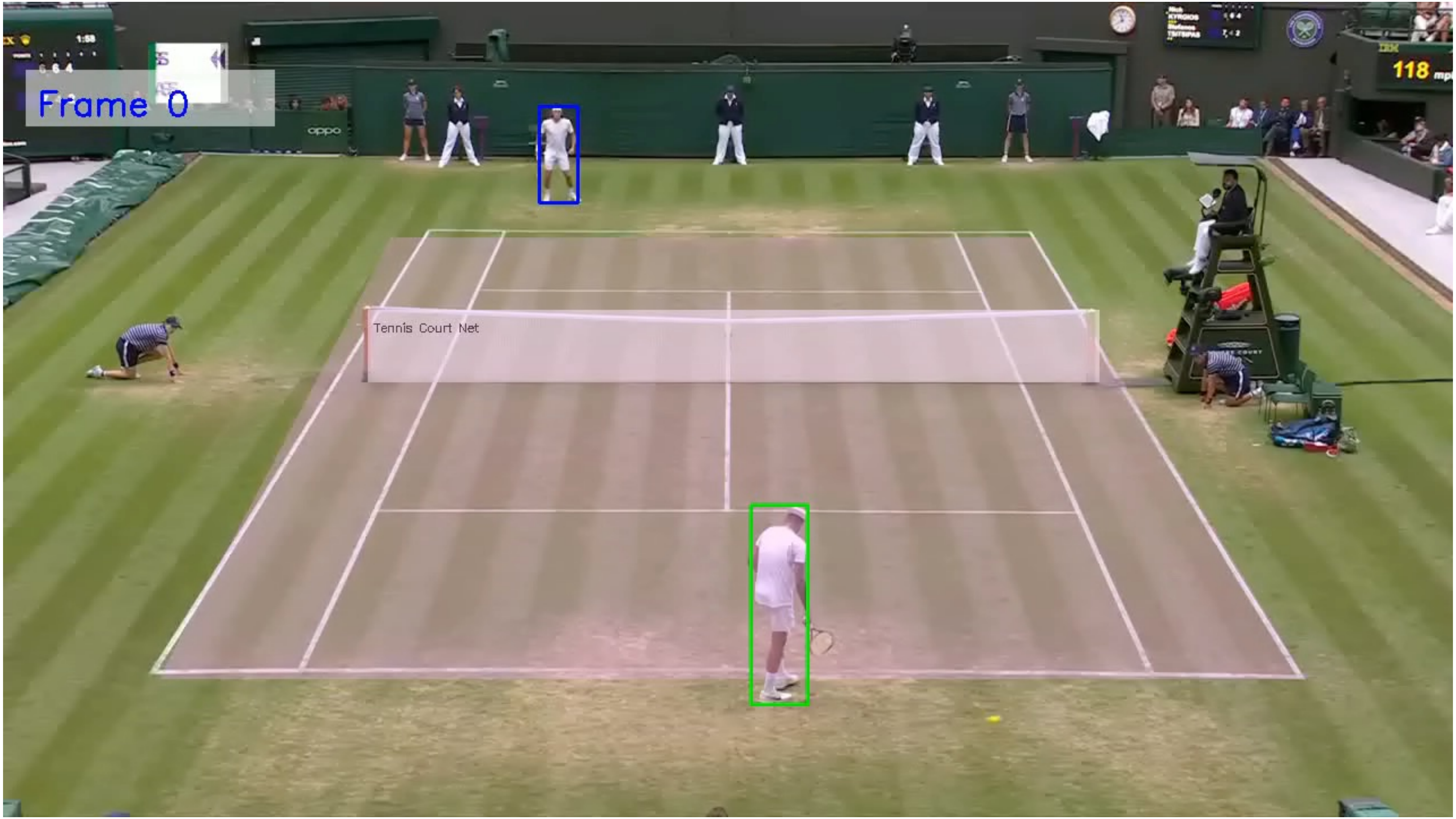
and pose/keypoint estimation

Ball Tracking

Action Recognition

Serve, Forehand, Volley, ...

Shot Directionality


Frame 0
Tennis Court Net
oppo
118

Metadata from other sensors and modalities

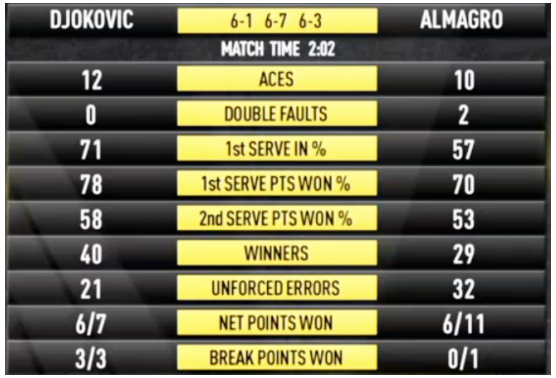
Score information

Ball speed (Radar)

Audio and vision excitement scores

Player information

Rally Length



DJOKOVIC	6-1 6-7 6-3	ALMAGRO
	MATCH TIME 2:02	
12	ACES	10
0	DOUBLE FAULTS	2
71	1st SERVE IN %	57
78	1st SERVE PTS WON %	70
58	2nd SERVE PTS WON %	53
40	WINNERS	29
21	UNFORCED ERRORS	32
6/7	NET POINTS WON	6/11
3/3	BREAK POINTS WON	0/1

Match analysis

Winner vs Error

Metadata: Computer vision + other modalities (input to large language model)

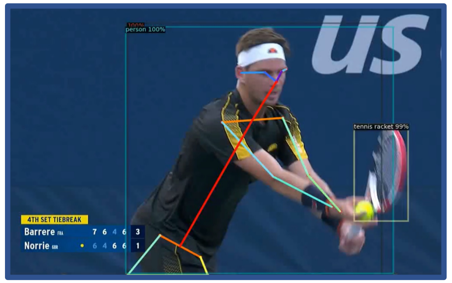
[{"id": "1127_10_db", "p1": "R. Peniston", "p2": "H. Laaksonen", "p1b": "", "p2b": "", "server": "R. Peniston", "set_num": 1, "point_number": 10, "points": "30-15", "game": "1-0", "description": "Peniston loses the point with a forehand volley unforced error.", "excitement_overall_score": 0.19817530137293593, "excitement_crowd_score": 0.2138548544686001, "excitement_analytics_score": 0, "excitement_gesture_score": 0.015000396221818044, "winner_of_point": 2, "high_excitement": 0, "long_rally": 0, "very_long_rally": 0, "winner": 0, "break_point": 0, "break_point_won": 0, "forced_error": 0, "ace": 0, "replay": 0, "volley": 1, "smash": 0, "net_approach": 1, "serve_number": 1, "file_name": "1127_10_db.mp4", "event": "wimbledon", "year": "2022", "server_id": 1, "event_type": "tennis", "day_msg": "Day 2: Tue, Jun 29", "court": "Court 12", "p1_fn": "Ryan Peniston", "p2_fn": "Henri Laaksonen", "p1b_fn": "", "p2b_fn": "", "p1_nat": "GBR", "p2_nat": "SUI", "p1b_nat": "", "p2b_nat": "", "Speed_MPH": 105, "DoubleFault": 0, "UnforcedError": 1, "NetPoint": 1, "BreakPoint": 0, "BreakPointWon": 0, "ServeDepth": "NCTL", "ReturnDepth": "ND", "rally_count": 4, "ServeWidth": "BW", "WinnerType": "0", "WinnerShotType": "0", "shots": {"top": "H. Laaksonen", "bottom": "R. Peniston", "seq": [{"who": "R. Peniston", "pos": "bottom", "stf": 58, "eff": 87, "act": "Serve", "where": "out wide"}, {"who": "H. Laaksonen", "pos": "top", "stf": 95, "eff": 126, "act": "Backhand", "where": "up the middle"}, {"who": "R. Peniston", "pos": "bottom", "stf": 142, "eff": 162, "act": "Backhand", "where": "out wide"}, {"who": "H. Laaksonen", "pos": "top", "stf": 176, "eff": 199, "act": "Backhand", "where": ""}, {"who": "R. Peniston", "pos": "bottom", "stf": 198, "eff": 254, "act": "Volley", "where": "up the middle"}, {"who": "R. Peniston", "pos": "bottom", "stf": 275, "eff": 292, "act": "NotAShot", "where": ""}]}, "pbp_file_path": "/dccstor/leonidka1/sonia_space/resultsOfMassiveProcessing/videoCVgeneratedMetadata/1127_10_db_metadata.json"}, {"id": "1137_150_db", "p1": "A. De Minaur", "p2": "H. Dellien", "p1b": "", "p2b": "", "server": "H. Dellien", "set_num": 3, "point_number": 150, "points": "15-30", "game": "4-4", "description": "De Minaur loses the point with a forehand forced error.", "excitement_overall_score": 0.40437613307910225, "excitement_crowd_score": 0.6252508636363142, "excitement_analytics_score": 0.1, "excitement_gesture_score": 0.015000289342755619, "winner_of_point": 2, "high_excitement": 0, "long_rally": 0, "very_long_rally": 0, "winner": 0, "break_point": 0, "break_point_won": 0, "forced_error": 1, "ace": 0, "replay": 0, "volley": 0, "smash": 0, "net_approach": 0, "serve_number": 1, "file_name": "1137_150_db.mp4", "event": "wimbledon", "year": "2022", "server_id": 2, "event_type": "tennis", "day_msg": "Day 2: Tue, Jun 29", "court": "No.1 Court", "p1_fn": "Alex De Minaur", "p2_fn": "Hugo Dellien", "p1b_fn": "", "p2b_fn": "", "p1_nat": "AUS", "p2_nat": "BOL", "p1b_nat": "", "p2b_nat": "", "Speed_MPH": 118, "DoubleFault": 0, "UnforcedError": 0, "NetPoint": 0, "BreakPoint": 0, "BreakPointWon": 0, "ServeDepth": "", "ReturnDepth": "", "rally_count": 1, "ServeWidth": "", "WinnerType": "0", "WinnerShotType": "0", "shots": {"top": "H. Dellien", "bottom": "A. De Minaur", "seq": [{"who": "H. Dellien", "pos": "top", "stf": 0, "eff": 32, "act": "Serve", "where": "out wide"}, {"who": "A. De Minaur", "pos": "bottom", "stf": 40, "eff": 45, "act": "Forehand", "where": "up the middle"}, {"who": "H. Dellien", "pos": "top", "stf": 55, "eff": 84, "act": "NotAShot", "where": ""}]}, "pbp_file_path": "/dccstor/leonidka1/sonia_space/resultsOfMassiveProcessing/videoCVgeneratedMetadata/1137_150_db_metadata.json"}, {"id": "1221_97_db", "p1": "D. Shapovalov", "p2": "B. Nakashima", "p1b": "", "p2b": "", "server": "B. Nakashima", "set_num": 3, "point_number": 97, "points": "15-0", "game": "0-0", "description": "Shapovalov wins the point with a forehand winner.", "excitement_overall_score": 0.6488774868315741, "excitement_crowd_score": 0.7745454437580231, "excitement_analytics_score": 0.55, "excitement_gesture_score": 0, "winner_of_point": 1, "high_excitement": 1, "long_rally": 1, "very_long_rally": 0, "winner": 1, "break_point": 0, "break_point_won": 0, "forced_error": 0, "ace": 0, "replay": 0, "volley": 0, "smash": 0, "net_approach": 2, "serve_number": 2, "file_name": "1221_97_db.mp4", "event": "wimbledon", "year": "2022", "server_id": 2, "event_type": "tennis", "day_msg": "Day 4: Thur, Jul 1", "court": "Court 12", "p1_fn": "Denis Shapovalov", "p2_fn": "Brandon Nakashima", "p1b_fn": "", "p2b_fn": "", "p1_nat": "CAN", "p2_nat": "USA", "p1b_nat": "", "p2b_nat": "", "Speed_MPH": 103, "DoubleFault": 0, "UnforcedError": 0, "NetPoint": 2, "BreakPoint": 0, "BreakPointWon": 0, "ServeDepth": "NCTL", "ReturnDepth": "ND", "rally_count": 6, "ServeWidth": "C", "WinnerType": "0", "WinnerShotType": "F", "shots": {"top": "D. Shapovalov", "bottom": "B. Nakashima", "seq": [{"who": "B. Nakashima", "pos": "bottom", "stf": 2, "eff": 36, "act": "Serve", "where": "up the middle"}, {"who": "D. Shapovalov", "pos": "top", "stf": 48, "eff": 70, "act": "Backhand", "where": "up the middle"}, {"who": "B. Nakashima", "pos": "bottom", "stf": 74, "eff": 95, "act": "Forehand", "where": "out wide"}, {"who": "D. Shapovalov", "pos": "top", "stf": 109, "eff": 120, "act": "UnknownShot", "where": ""}, {"who": "B. Nakashima", "pos": "bottom", "stf": 132, "eff": 154, "act": "Volley", "where": "up the middle"}, {"who": "D. Shapovalov", "pos": "top", "stf": 171, "eff": 193, "act": "UnknownShot", "where": ""}, {"who": "B. Nakashima", "pos": "bottom", "stf": 192, "eff": 203, "act": "Volley", "where": ""}]}, "pbp_file_path": "/dccstor/leonidka1/sonia_space/resultsOfMassiveProcessing/videoCVgeneratedMetadata/1221_97_db_metadata.json"}, {"id": "1204_229_db", "p1": "T. Van Rijthoven", "p2": "R. Opelka", "p1b": "", "p2b": "", "server": "T. Van Rijthoven", "set_num": 4, "point_number": 229, "points": "40-40", "game": "0-1", "description": "Opelka loses the point with a forehand forced error.", "excitement_overall_score": 0.2833576388458572, "excitement_crowd_score": 0.47660311447231984, "excitement_analytics_score": 0.1, "excitement_gesture_score": 0.01500647568195426, "winner_of_point": 1, "high_excitement": 0, "long_rally": 0, "very_long_rally": 0, "winner": 0, "break_point": 0, "break_point_won": 0, "forced_error": 1, "ace": 0, "replay": 0, "volley": 0, "smash": 0, "net_approach": 0, "serve_number": 1, "file_name": "1204_229_db.mp4", "event": "wimbledon", "year": "2022", "server_id": 1, "event_type": "tennis", "day_msg": "Day 3: Wed, Jun 30", "court": "Court 4", "p1_fn": "Tim Van Rijthoven", "p2_fn": "Reilly Opelka", "p1b_fn": "", "p2b_fn": "", "p1_nat": "NED", "p2_nat": "USA", "p1b_nat": "", "p2b_nat": "", "Speed_MPH": 0, "DoubleFault": 0, "UnforcedError": 0, "NetPoint": 0, "BreakPoint": 0, "BreakPointWon": 0, "ServeDepth": "", "ReturnDepth": "", "rally_count": 0, "ServeWidth": "", "WinnerType": "0", "WinnerShotType": "0", "shots": {"top": "R. Opelka", "bottom": "T. Van Rijthoven", "seq": [{"who": "T. Van Rijthoven", "pos": "bottom", "stf": 12, "eff": 41, "act": "Serve", "where": "up the middle"}, {"who": "R. Opelka", "pos": "top", "stf": 42, "eff": 82, "act": "Forehand", "where": "up the middle"}, {"who": "T. Van Rijthoven", "pos": "bottom", "stf": 87, "eff": 102, "act": "Forehand", "where": "up the middle"}, {"who": "R. Opelka", "pos": "top", "stf": 130, "eff": 184, "act": "Backhand", "where": "out wide"}, {"who": "T. Van Rijthoven", "pos": "bottom", "stf": 196, "eff": 203, "act": "Backhand", "where": "down the line"}, {"who": "R. Opelka", "pos": "top", "stf": 202, "eff": 214, "act": "NotAShot", "where": ""}]}, "pbp_file_path": "/dccstor/leonidka1/sonia_space /resultsOfMassiveProcessing/videoCVgeneratedMetadata/1204_229_db_metadata.json"}, {"id": "1158_219_db", "p1": "J. Sousa", "p2": "R. Gasquet", "p1b": "", "p2b": "", "server": "R. Gasquet", "set_num": 4, "point_number": 219, "points": "40-AD", "game": "1-1", "description": "Gasquet loses the point with a double fault.", "excitement_overall_score": 0.12421195070662616, "excitement_crowd_score": 0.13604091128807183, "excitement_analytics_score": 0, "excitement_gesture_score": 0.015028462476170608, "winner_of_point": 1, "high_excitement": 0, "long_rally": 0, "very_long_rally": 0, "winner": 0, "break_point": 0, "break_point_won": 0, "forced_error": 0, "ace": 0, "replay": 0, "volley": 0, "smash": 0, "net_approach": 0, "serve_number": 0, "file_name": "1158_219_db.mp4", "event": "wimbledon", "year": "2022", "server_id": 2, "event_type": "tennis", "day_msg": "Day 2: Tue, Jun 29", "court": "Court 15", "p1_fn": "Joao Sousa", "p2_fn": "Richard Gasquet", "p1b_fn": "", "p2b_fn": "", "p1_nat": "POR", "p2_nat": "FRA", "p1b_nat": "", "p2b_nat": "", "Speed_MPH": 0, "DoubleFault": 2, "UnforcedError": 2, "NetPoint": 0, "BreakPoint": 0, "BreakPointWon": 0, "ServeDepth": "CTL", "ReturnDepth": "", "rally_count": 0, "ServeWidth": "BW", "WinnerType": "0", "WinnerShotType": "0", "shots": {"top": "R. Gasquet", "bottom": "J. Sousa", "seq": [{"who": "R. Gasquet", "pos": "top", "stf": 35, "eff": 62, "act": "Serve", "where": "up the middle"}, {"who": "R. Gasquet", "pos": "top", "stf": 75, "eff": 82, "act": "UnknownShot", "where": ""}, {"who": "R. Gasquet", "pos": "top", "stf": 199, "eff": 271, "act": "Serve", "where": ""}, {"who": "J. Sousa", "pos": "bottom", "stf": 279, "eff": 291, "act": "Backhand", "where": "up the middle"}, {"who": "R. Gasquet", "pos": "top", "stf": 298, "eff": 320, "act": "NotAShot", "where": ""}, {"who": "R. Gasquet", "pos": "top", "stf": 413, "eff": 434, "act": "NotAShot", "where": ""}]}, "pbp_file_path": "/dccstor/leonidka1/sonia_space/resultsOfMassiveProcessing/videoCVgeneratedMetadata/1158_219_db_metadata.json"}, {"id": "1151_134_db", "p1": "D. Novak", "p2": "F. Bagnis", "p1b": "", "p2b": "", "server": "F. Bagnis", "set_num": 2, "point_number": 134, "points": "40-40", "game": "2-3", "description": "Bagnis wins the point with a backhand volley winner.", "excitement_overall_score": 0.4132240626662806, "excitement_crowd_score": 0.40765417486424604, "excitement_analytics_score": 0.3, "excitement_gesture_score": 0, "winner_of_point": 2, "high_excitement": 0, "long_rally": 0, "very_long_rally": 0, "winner": 2, "break_point": 0, "break_point_won": 0, "forced_error": 0, "ace": 0, "replay": 0, "volley": 1, "smash": 0, "net_approach": 3, "serve_number": 1, "file_name": "1151_134_db.mp4", "event": "wimbledon", "year": "2022", "server_id": 2, "event_type": "tennis", "day_msg": "Day 2: Tue, Jun 29", "court": "Court 6", "p1_fn": "Dennis Novak", "p2_fn": "Facundo Bagnis", "p1b_fn": "", "p2b_fn": "", "p1_nat": "AUT", "p2_nat": "ARG", "p1b_nat": "", "p2b_nat": "", "Speed_MPH": 0, "DoubleFault": 0, "UnforcedError": 0, "NetPoint": 3, "BreakPoint": 0, "BreakPointWon": 0, "ServeDepth": "NCTL", "ReturnDepth": "ND", "rally_count": 0, "ServeWidth": "B", "WinnerType": "0", "WinnerShotType": "B", "shots": {"top": "D. Novak", "bottom": "F. Bagnis", "seq": [{"who": "F. Bagnis", "pos": "bottom", "stf": 9, "eff": 13, "act": "Serve", "where": "up the middle"}, {"who": "D. Novak", "pos": "top", "stf": 20, "eff": 47, "act": "Backhand", "where": "up the middle"}, {"who": "F. Bagnis", "pos": "bottom", "stf": 50, "eff": 65, "act": "Forehand", "where": "up the middle"}, {"who": "D. Novak", "pos": "top", "stf": 72, "eff": 126, "act": "Forehand", "where": ""}, {"who": "F. Bagnis", "pos": "bottom", "stf": 125, "eff": 133, "act": "Volley", "where": ""}, {"who": "D. Novak", "pos": "top", "stf": 132, "eff": 135, "act": "UnknownShot", "where": ""}, {"who": "F. Bagnis", "pos": "bottom", "stf": 134, "eff": 146, "act": "Volley", "where": "out wide"}, {"who": "D. Novak", "pos": "top", "stf": 146, "eff": 154, "act": "NotAShot", "where": ""}, {"who": "F. Bagnis", "pos": "bottom", "stf": 153, "eff": 155, "act": "NotAShot", "where": ""}, {"who": "D. Novak", "pos": "top", "stf": 154, "eff": 167, "act": "NotAShot", "where": ""}]}, "pbp_file_path": "/dccstor/leonidka1/sonia_space/resultsOfMassiveProcessing/videoCVgeneratedMetadata/1151_134_db_metadata.json"}, {"id": "1105_210_db", "p1": "N. Basilashvili", "p2": "L. Rosol", "p1b": "", "p2b": "", "server": "N. Basilashvili", "set_num": 4, "point_number": 210, "points": "0-0", "game": "null-null", "description": "Rosol loses the point with a forehand forced error.", "excitement_overall_score": 0.16482244279582572, "excitement_crowd_score": 0.25398993260619507, "excitement_analytics_score": 0.1, "excitement_gesture_score": 0.015006083878708913, "winner_of_point": 1, "high_excitement": 0, "long_rally": 0, "very_long_rally": 0, "winner": 0, "break_point": 0, "break_point_won": 0, "forced_error": 1, "ace": 0, "replay": 0, "volley": 0, "smash": 0, "net_approach": 0, "serve_number": 2, "file_name": "1105_210_db.mp4", "event": "wimbledon", "year": "2022", "server_id": 1, "event_type": "tennis", "day_msg": "Day 1: Mon, Jun 28", "court": "Court 15", "p1_fn": "Nikoloz Basilashvili", "p2_fn": "Lukas Rosol", "p1b_fn": "", "p2b_fn": "", "p1_nat": "GEO", "p2_nat": "CZE", "p1b_nat": "", "p2b_nat": "", "Speed_MPH": 0, "DoubleFault": 0, "UnforcedError": 0, "NetPoint": 0, "BreakPoint": 0, "BreakPointWon": 0, "ServeDepth": "", "ReturnDepth": "", "rally_count": 0, "ServeWidth": "", "WinnerType": "0", "WinnerShotType": "0", "shots": {"top": "N. Basilashvili", "bottom": "L. Rosol", "seq": [{"who": "N. Basilashvili", "pos": "top", "stf": 12, "eff": 26, "act": "Serve", "where": "up the middle"}, {"who": "L. Rosol", "pos": "bottom", "stf": 29, "eff": 61, "act": "Backhand", "where": ""}, {"who": "N. Basilashvili", "pos": "top", "stf": 97, "eff": 265, "act": "Serve", "where": "up the middle"}, {"who": "L. Rosol", "pos": "bottom", "stf": 265, "eff": 275, "act": "Forehand", "where": ""}, {"who": "N. Basilashvili", "pos": "top", "stf": 290, "eff": 296, "act": "NotAShot", "where": ""}]}, "pbp_file_path": "/dccstor/leonidka1/sonia_space/resultsOfMassiveProcessing/videoCVgeneratedMetadata /1105_210_db_metadata.json"}, {"id": "1115_171_db", "p1": "T. Griekspoor", "p2": "F. Fognini", "p1b": "", "p2b": "", "server": "F. Fognini", "set_num": 3, "point_number": 171, "points": "15-40", "game": "5-2", "description": "Fognini wins the point with an ace.", "excitement_overall_score": 0.3196034629821145, "excitement_crowd_score": 0.5567825983088074, "excitement_analytics_score": 0.1, "excitement_gesture_score": 0.015079165274028783, "winner_of_point": 2, "high_excitement": 0, "long_rally": 0, "very_long_rally": 0, "winner": 2, "break_point": 0, "break_point_won": 0, "forced_error": 0, "ace": 2, "replay": 0, "volley": 0, "smash": 0, "net_approach": 0, "serve_number": 2, "file_name": "1115_171_db.mp4", "event": "wimbledon", "year": "2022", "server_id": 2, "event_type": "tennis", "day_msg": "Day 1: Mon, Jun 28", "court": "Court 16", "p1_fn": "Tallon Griekspoor", "p2_fn": "Fabio Fognini", "p1b_fn": "", "p2b_fn": "", "p1_nat": "NED", "p2_nat": "ITA", "p1b_nat": "", "p2b_nat": "", "Speed_MPH": 0, "DoubleFault": 0, "UnforcedError": 0, "NetPoint": 0, "BreakPoint": 0, "BreakPointWon": 0, "ServeDepth": "", "ReturnDepth": "", "rally_count": 0, "ServeWidth": "", "WinnerType": "0", "WinnerShotType": "0", "shots": {"top": "T. Griekspoor", "bottom": "F. Fognini", "seq": [{"who": "F. Fognini", "pos": "bottom", "stf": 19, "eff": 45, "act": "Serve", "where": "up the middle"}, {"who": "T. Griekspoor", "pos": "top", "stf": 55, "eff": 92, "act": "Backhand", "where": "up the middle"}, {"who": "F. Fognini", "pos": "bottom", "stf": 127, "eff": 239, "act": "Serve", "where": "up the middle"}, {"who": "T. Griekspoor", "pos": "top", "stf": 253, "eff": 305, "act": "NotAShot", "where": ""}]}, "pbp_file_path": "/dccstor/leonidka1/sonia_space/resultsOfMassiveProcessing/videoCVgeneratedMetadata/1115_171_db_metadata.json"}, {"id": "1144_109_db", "p1": "A. Balazs", "p2": "R. Bautista Agut", "p1b": "", "p2b": "", "server": "R. Bautista Agut", "set_num": 3, "point_number": 109, "points": "15-15", "game": "2-1", "description": "Balazs loses the point with a backhand forced error.", "excitement_overall_score": 0.18151347945402357, "excitement_crowd_score": 0.23575143355335748, "excitement_analytics_score": 0.1, "excitement_gesture_score": 0.015001179158329013, "winner_of_point": 2, "high_excitement": 0, "long_rally": 0, "very_long_rally": 0, "winner": 0, "break_point": 0, "break_point_won": 0, "forced_error": 1, "ace": 0, "replay": 0, "volley": 0, "smash": 0, "net_approach": 0, "serve_number": 1, "file_name": "1144_109_db.mp4", "event": "wimbledon", "year": "2022", "server_id": 2, "event_type": "tennis", "day_msg": "Day 2: Tue, Jun 29", "court": "Court 9", "p1_fn": "Attila Balazs", "p2_fn": "Roberto Bautista Agut", "p1b_fn": "", "p2b_fn": "", "p1_nat": "HUN", "p2_nat": "ESP", "p1b_nat": "", "p2b_nat": "", "Speed_MPH": 0, "DoubleFault": 0, "UnforcedError": 0, "NetPoint": 0, "BreakPoint": 0, "BreakPointWon": 0, "ServeDepth": "", "ReturnDepth": "", "rally_count": 0, "ServeWidth": "", "WinnerType": "0", "WinnerShotType": "0", "shots": {"top": "A. Balazs", "bottom": "R. Bautista Agut", "seq": [{"who": "R. Bautista Agut", "pos": "bottom", "stf": 10, "eff": 21, "act": "NotAShot", "where": ""}, {"who": "A. Balazs", "pos": "top", "stf": 20, "eff": 28, "act": "NotAShot", "where": ""}, {"who": "R. Bautista Agut", "pos": "bottom", "stf": 35, "eff": 39, "act": "NotAShot", "where": ""}, {"who": "A. Balazs", "pos": "top", "stf": 40, "eff": 47, "act": "NotAShot", "where": ""}, {"who": "R. Bautista Agut", "pos": "bottom", "stf": 46, "eff": 87, "act": "NotAShot", "where": ""}, {"who": "R. Bautista Agut", "pos": "bottom", "stf": 108, "eff": 124, "act": "Serve", "where": "up the middle"}, {"who": "A. Balazs", "pos": "top", "stf": 133, "eff": 140, "act": "NotAShot", "where": ""}]}, "pbp_file_path": "/dccstor/leonidka1/sonia_space/resultsOfMassiveProcessing/videoCVgeneratedMetadata/1144_109_db_metadata.json"}, {"id": "1219_21_db", "p1": "A. De Minaur", "p2": "J. Draper", "p1b": "", "p2b": "", "server": "J. Draper", "set_num": 1, "point_number": 21, "points": "15-40", "game": "1-2", "description": "Draper wins the point with an ace.", "excitement_overall_score": 0.3127906036660247, "excitement_crowd_score": 0.2544923735988407, "excitement_analytics_score": 0.1, "excitement_gesture_score": 0.015133160588898463, "winner_of_point": 2, "high_excitement": 0, "long_rally": 0, "very_long_rally": 0, "winner": 2, "break_point": 0, "break_point_won": 0, "forced_error": 0, "ace": 2, "replay": 0, "volley": 0, "smash": 0, "net_approach": 0, "serve_number": 1, "file_name": "1219_21_db.mp4", "event": "wimbledon", "year": "2022", "server_id": 2, "event_type": "tennis", "day_msg": "Day 4: Thur, Jul 1", "court": "No.1 Court", "p1_fn": "Alex De Minaur", "p2_fn": "Jack Draper", "p1b_fn": "", "p2b_fn": "", "p1_nat": "AUS", "p2_nat": "GBR", "p1b_nat": "", "p2b_nat": "", "Speed_MPH": 131, "DoubleFault": 0, "UnforcedError": 0, "NetPoint": 0, "BreakPoint": 0, "BreakPointWon": 0, "ServeDepth": "CTL", "ReturnDepth": "", "rally_count": 1, "ServeWidth": "C", "WinnerType": "0", "WinnerShotType": "0", "shots": {"top": "A. De Minaur", "bottom": "J. Draper", "seq": [{"who": "J. Draper", "pos": "bottom", "stf": 0, "eff": 25, "act": "Serve", "where": "up the middle"}, {"who": "A. De Minaur", "pos": "top", "stf": 37, "eff": 63, "act": "Serve", "where": ""}]}, "pbp_file_path": "/dccstor/leonidka1/sonia_space/resultsOfMassiveProcessing/videoCVgeneratedMetadata/1219_21_db_metadata.json"}, {"id": "1107_106_db", "p1": "T. Van Rijthoven", "p2": "F. Delbonis", "p1b": "", "p2b": "", "server": "F. Delbonis", "set_num": 2, "point_number": 106,

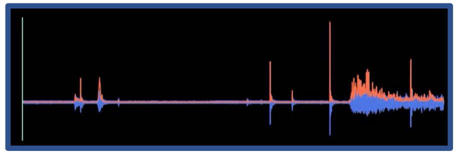
Turning multi-sensory data into fluent commentary using Large Language Models

After a wide out serve by Peniston, Laaksonen returns the service with a short backhand. Peniston then approaches the net, hits a forehand, but Laaksonen strikes back with a magnificent backhand. What a beautiful shot! Peniston makes all efforts to save the point but hits the net with a volley unforced error. The game is now tied: 30-30

Billion-scale IBM Language Model

- IBM Sandstone.3B model
- Pretrained on petabytes of data
- Fine-tuned on real and synthetic commentary data





Video

Sound

Match Data

Excitement Scores

Projection into a unified embedding space
(v1: Data to Text)

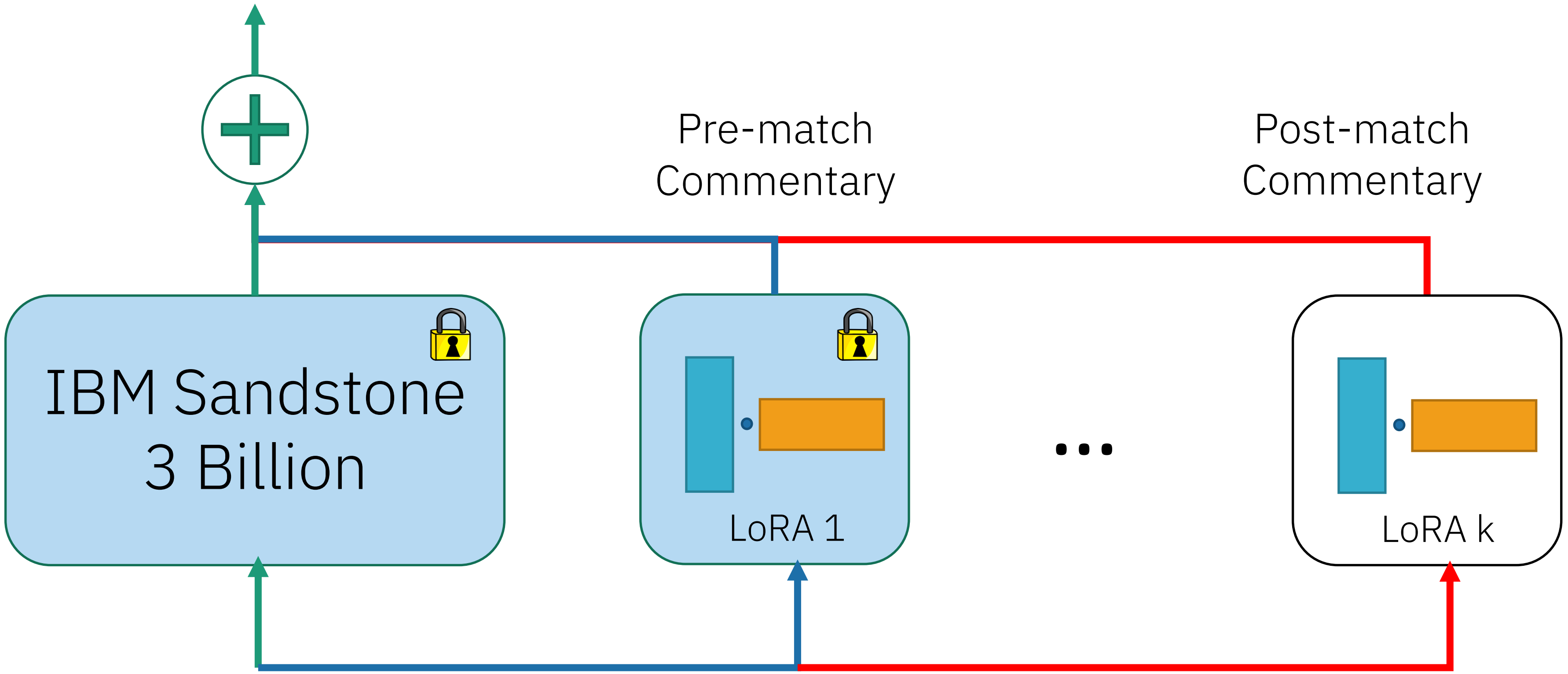
IBM Sandstone 3B model

- Encoder-Decoder Architecture
- Base model pre-trained on petabytes of data
- Long context modeling with ALiBi

- Tuned for AI commentary using Synthetic and Real Data
- Efficient tuning is achieved using Layered LoRA (next)

Our Layered LoRA Architecture (LaLo)

[James Smith et al, CVPR 2023]

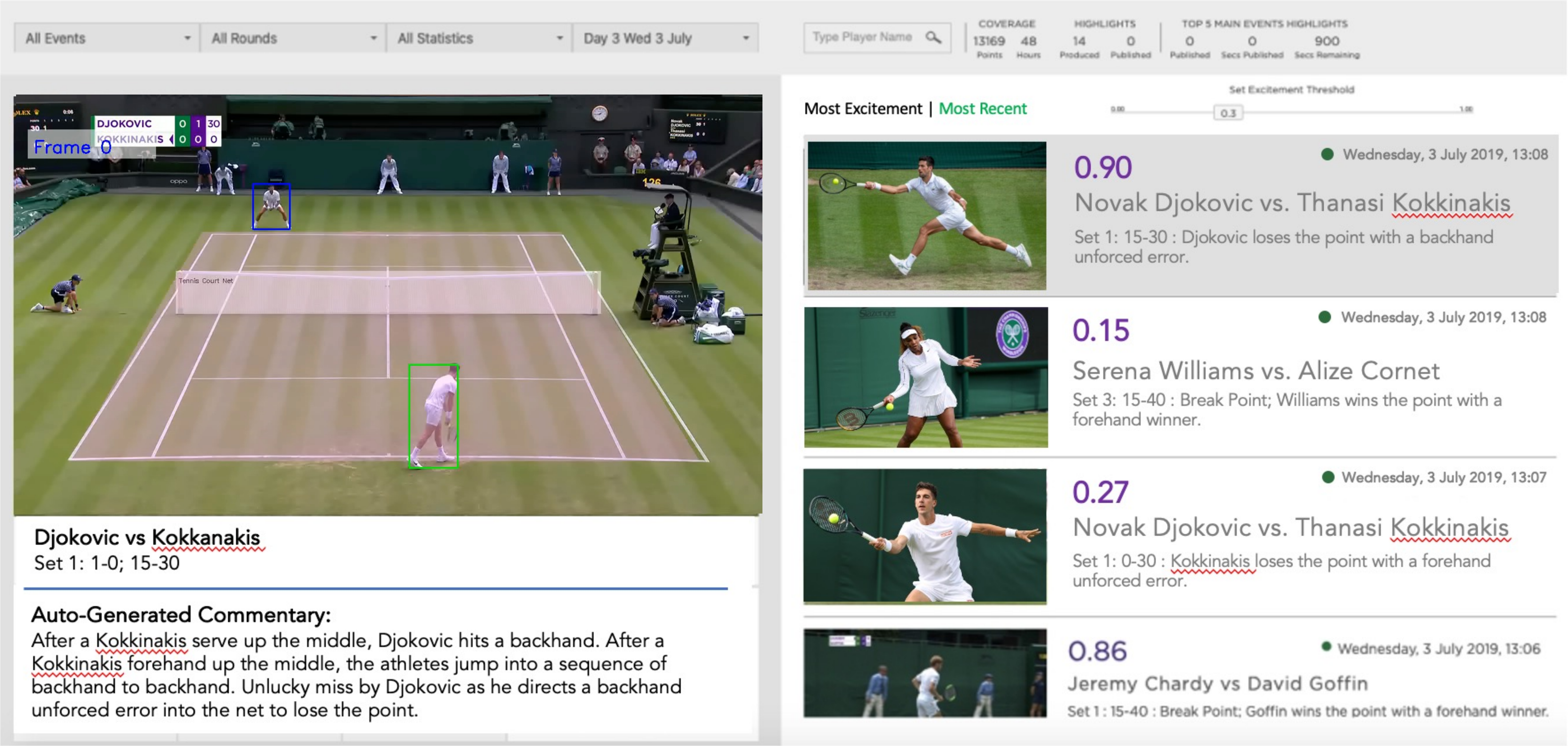


LaLo can support task modularity & efficient recall of past models

Efficient adaptation: < 3% of all parameters are adapted !

Wimbledon Multimodal Perception and Commentary

IBM



Conclusions

- Era of Foundation Models: Training with billions of images (vision) and trillions of tokens (language)
- Take-away message: focus on data quality (instead of just data quantity) for safer and more efficient models
- Learning with real + synthetic data is a promising way to achieve this goal
- Outlook: Synthetic data for understanding what makes for a good pre-training model

Thank you!

See more at http://rogerioferis.org