MOHAMED BIN ZAYED UNIVERSITY OF ARTIFICIAL INTELLIGENCE

GenAI Content Detection Task 1: English and Multilingual Machine-Generated Text Detection: AI vs. Human

Presenter: Yuxia Wang

Previous Shared Tasks on Machine-generated Text (MGT) Detection

- **English**
  - 2023 ALTA shared task (ChatGPT generated)
  - DAGPap22 shared task (Scientific papers)
  - SemEval 2024 shared task 8 (4 sub tasks)
- **Other languages**
  - RuATD Shared task 2022(Russian)
  - IberLEF 2023 (Spanish)
  - CLIN33 (Dutch)
  - SemEval-2024 Task 8 (9 languages)

COLING 2025 GenAI Shared Task 1

Overview

Task Description

Dataset

Baselines

Task 1 Overview

- **Goal:**
  Develop robust and generalized MGT detectors across languages and domains.
- **Binary classification:** human vs. machine
  Subtask A: Monolingual – English
  Subtask B: Multilingual – 15 languages in training and test sets, with 9 overlap
- **Participants:**
  Subtask A: Monolingual: 36 submissions
  Subtask B: Multilingual: 26 submissions
- **System description paper submissions:**
  18 papers were accepted

Task 1 Description

- **Timeline:**
  - **Development Phase**: Aug 27 – Oct 29, 2024
    - Labeled training/validation data provided.
    - Unlabeled dev-test set for generalization testing.
  - **Test Phase**: Oct 30 – Nov 4, 2024
    - Test subset texts provided with limited submission attempts.
    - Dev-test labels revealed.
  - **Paper Submission Phase**: Nov 21 – Dec 13, 2024

- **Post-Test Analysis:**
  - Test set labels released for ablation studies.
- **Rules:**
  - Use only organizer-provided data for model development.
  - External training data strictly prohibited.

Dataset – Subtask A: Monolingual English

Split	Source	Data License	#Generators	#Domains	Human	MGT	H+M	Total
Train	HC3	CC BY-SA-4.0	1	5	39,140	18,671	57,811	610,767
	M4GT	CC BY-SA-4.0	14	6	86,782	181,081	267,863	
	MAGE	Apache-2.0	27	14	103,000	182,093	285,093	
Dev	HC3	CC BY-SA-4.0	1	5	16,855	7,917	24,772	261,758
	M4GT	CC BY-SA-4.0	14	6	37,220	77,267	114,487	
	MAGE	Apache-2.0	27	14	44,253	78,246	122,499	
Dev-test	RAID	MIT	0	–	13,371	0	13,371	32,557
	LLM-DetectAIve	CC BY-SA-4.0	5	–	0	19,186	19,186	
Test	CUDRT	CC BY-SA-4.0	6	6	12,287	10,691	22,978	73,941
	IELTS	Apache-2.0	2	1	11,382	13,318	24,700	
	NLPeer	Apache-2.0	1	1	5,326	5,376	10,702	
	PeerSum	Apache-2.0	2	1	5,080	6,995	12,075	
	MixSet	CC BY-SA-4.0	7	9	600	2,886	3,486	
Total					375,296	603,727	979,023	**979,023**

Dataset – Subtask A: Monolingual English Test Set Distribution

Source / Domain	License	# Human	# MGT	LLM Generator List
CUDRT-en subset	CC BY-SA 4.0	12939	10800	GPT-3.5-turbo, Llama2, Llama3, ChatGLM, Baichuan, Qwen (1800 samples each)
Mixset	CC BY-SA 4.0	600	3000	-
LLM-DetectAIve-IELTS	huggingface	1635	900	llama-3.1-70B-versatile (900 samples)
IELTSDuck	Apache-2.0	10932	12418	GPT-4o-mini-2024-07-18, (10932), llama-3.1-70B-versatile (1486)
NLPeer	Apache-2.0	5376	5376	GPT-4o-2024-05-13 (5376)
Peersum	Github	5157	6997	GPT-4o-2024-08-06 (3501), GPT-4o-mini-2024-07-18 (3496)
Total	-	36639	39491	-
After deduplication	-	35393	39363	-
After removing short text	-	34675	39266	-

Dataset – Subtask B: Multilingual

Split	Source	Data License	Lang	#Generators	#Domains	Human	MGT	H+M	Total
Train	HC3	CC BY-SA-4.0	zh, en	1	9	54,655	30,670	85,325	674,083
	M4GT	CC BY-SA-4.0	9	16	13	100,359	203,525	303,884	
	MAGE	Apache-2.0	en	27	14	102,954	181,920	284,874	
Dev	HC3	CC BY-SA-4.0	zh, en	1	9	22,981	12,718	35,699	288,894
	M4GT	CC BY-SA-4.0	9	16	13	42,886	87,591	130,477	
	MAGE	Apache-2.0	en	27	14	44,299	78,419	122,718	
Dev-test	MULTITuDE	GPL-3.0	11	8	–	7,992	66,089	74,081	74,081
Test	29 sources	–	15	19	–	73,634	77,791	151,425	151,425
Total						449,760	738,723	1,188,483	**1,188,483**

Train and Development 15 languages: Arabic, **Bulgarian**, **Catalan**, Chinese, **Czech**, Dutch, English, German, Indonesian, Italian, **Portuguese**, Russian, **Spanish**, **Ukrainian**, Urdu.

Test 15 languages: Arabic, Chinese, Dutch, German, **Hebrew, Hindi**, Indonesian, Italian, **Japanese, Kazakh, Norwegian**, Russian, Spanish, Urdu, and **Vietnamese.**

Dataset – Subtask B: Multilingual Test Set Distribution (1)

Source / Domain	Language	# Human	# MGT	LLM Generator List
Cudrt-Subset	Chinese	12565	1500	GPT-3.5 (300), Qwen (300), GPT-4 (300), ChatGLM (300), Baichuan (300)
High School Student Essay	Chinese	3502	1556	GLM-4-9b-chat (778), Claude-3.5-sonnet (778)
Zhihu-Qa	Chinese	12524	10269	GPT-4o-2024-08-06 (3423), GPT-4o-mini-2024-07-18 (6846)
Mnbvc-Qa-Zhihu	Chinese	3000	3000	GPT-4o-2024-05-13 (3000)
Govreport	Chinese	2975	17695	GPT-4o-2024-05-13 (5932), ChatGLM3-6B (5821)
Easc (Summary)	Arabic	153	306	GPT-4o-2024-08-06 (306)
Tweets	Arabic	1400	3400	GPT-4 (1700), GPT-4o-2024-08-06 (1400), Qwen-2.5 72B (300)
Kalimat Youm 7 News	Arabic	1000	2000	GPT-4o-2024-05-13 (1000), Ace-GPT (1000)
Sanad (News)	Arabic	3000	3000	GPT-4o-2024-05-13 (3000)
Summaries	Russian	6562	6582	GPT-4o-2024-08-06 (3300), Vikhrmodels/Vikhr-Nemo-12B-Instruct-R-21-09-24 (3282)
News	Russian	6494	6539	GPT-4o-2024-08-06 (3295), Vikhrmodels/Vikhr-Nemo-12B-Instruct-R-21-09-24 (3244)
Wikipedia	Russian	1025	3049	GPT-4-0613 (999), Vikhrmodels/it-5.4-fp16-orpo-v2 (1025), AnatoliiPotapov/T-lite-instruct-0.1 (1025)

Dataset – Subtask B: Multilingual Test Set Distribution (2)

Source / Domain	Language	# Human	# MGT	LLM Generator List
Wikipedia	Hebrew	1182	2173	GPT-4-0613 (991), dicta-il/dictalm2.0-instruct (1182)
Wikipedia	German	1865	2529	GPT-4-0613 (957), LeoLM/leo-hessianai-13b-chat (1572)
Wikipedia	Norwegian	1544	2543	GPT-4-0613 (999), norallm/normistral-7b-warm-instruct (1544)
Wikipedia	Spanish	600	600	Llama 3.1 405B instruct (600)
Wikipedia	Dutch	600	600	Llama 3.1 405B instruct (600)
Wikipedia	kaz	1300	1300	GPT-4o-2024-08-06 (1300)
Dice (News)	Italian	2800	2800	Llama 3.1 405B instruct (2800)
News	Urdu	13497	17472	GPT-4o-2024-08-06 (17472)
News	Hindi	600	600	GPT-4o-2024-08-06 (600)
News	Japanese	300	300	GPT-4o-2024-08-06 (300)
News	Vietnamese	600	600	GPT-4o-2024-08-06 (600)
Wikipedia	Vietnamese	600	600	GPT-4o-2024-08-06 (600)
Poetry	Indonesian	600	600	GPT-4o-2024-08-06 (600)
Total	-	80288	91613	-
Non-duplicated	-	78424	79305	-
Remove Short Text	-	73634	77791	-

Baselines

We fine-tuned pre-trained Transformer encoders on the training sets as baselines.

Subtask A: RoBERTa

Subtask B: XLM-R

Task	Set	Accuracy	F1
Subtask A	Dev	96.2	95.9 / 96.2
	Dev-Test	83.1	81.6 / 82.6
	Test	74.9	73.4 / 73.8
Subtask B	Dev	95.2	94.8 / 95.2
	Dev-Test	84.7	65.5 / 85.7
	Test	74.7	74.2 / 74.3

Baseline performance on the Dev, Dev-Test, and Test sets according to accuracy and macro/micro F1.

Participants

- Monolingual
- Multilingual

Subtask A: Overview

- Number of submissions: 36
- Highest scores: 1st – 83.1, 2nd – 83.0, 3rd – 82.3 (Macro F1)
- Most used methodologies:
  - Small LM: 10 submissions
  - Large LM: 6 submissions
  - Ensembling: 4 submissions
  - Feature Combination: 3 submissions

Subtask A: Top-3 Team Detection Approaches

1. **Advacheck**: Shared Transformer Encoder (DeBERTa-v3-base) with several classification heads, a binary classification head for MGT detection and multiclass heads for text domain classification
2. **Unibuc-NLP**: Finetuning both Masked Language Model (XLM-RoBERTa) and Causal Language Model (Qwen2.5B)
3. **Fraunhofer-SIT**: Combined MGT detection adapter with a multi-genre natural language inference adapter over RoBERTa-base.

Subtask B: Overview

- Number submissions: 26
- Highest scores: 1st – 79.16, 2nd – 75.57, 3rd – 75.32 (Macro F1)
- Most used methodologies:
  - Small LM: 5 submissions
  - Large LM: 3 submissions
  - Ensembling: 3 submissions
  - Feature Combination: 1 submission

Subtask B: Top-3 Team Detection Approaches

1. **Grape**: Finetuning small LMs and training an ensemble model on top of them.
2. **Nota AI**: Combining a language identification tool, finetuning a multilingual LM, and token-level probability distributions extracted from various LMs.
3. **Lux Veri**: Ensembling RemBERT, XLM-RoBERTa-base, and BERT-base-multilingual-cased using inverse pseudo-perplexity weighting.

Analysis

Monolingual

Multilingual

Subtask A: Analysis of Monolingual Performance

Overall Performance

- Generally, in-domain data performance > out-of-domain data performance

In-Domain Data Performance

PeerReview:
- Top systems (Rank 1-5) scored ≥ 80%, with highest at 89.9%; Consistently high performance (≥ 90% for all systems above baseline).
- Training on PeerRead (M4GT-Bench) enabled effective domain-specific pattern recognition.

IELTS Essays:
- Only top 5 systems achieved ≥ 80%.
- Performance impacted by subtle differences between training and test data (e.g., native vs. non-native English authors).

Out-of-Domain Dataset Performance

MixSet:
- Diverse genres (game reviews, emails, blogs, speech) led to performance drops (48–66.7%); Teams above baseline struggled (≤ 5% improvement), while lower-ranked teams achieved significant gains (up to 82.3%).
- Humanization and adaptation of machine-generated text (MGT) increased difficulty.

CUDRT
- Partial domain overlap with training data (e.g., news).
- Scores ranged 65–75%, reflecting moderate adaptability.

Rank	All	MixSet	CUDRT	IELTS	PeerReview
1	83.1	48.0	67.1	89.9	97.2
2	83.3	66.7	75.9	82.6	94.1
3	82.9	58.9	71.0	88.8	92.1
4	82.2	64.7	73.2	79.1	97.4
5	81.8	59.2	72.7	80.8	95.5
6	80.7	47.2	72.6	78.1	96.9
7	75.7	54.9	71.0	63.1	97.2
8	79.3	62.3	75.4	69.0	97.2
9	78.0	60.0	74.6	66.3	96.9
10	76.4	59.8	75.5	64.2	93.2
11	75.5	60.9	70.3	66.9	92.5
12	75.7	56.6	74.0	61.9	95.2
13	75.2	62.8	70.8	65.3	92.2
14	75.1	66.6	72.8	62.7	92.2
BL	74.9	62.0	72.1	63.4	92.2
15	74.8	73.2	71.9	63.0	90.8
-	73.2	53.5	71.3	62.8	89.3
16	73.9	64.3	71.2	62.6	90.3
17	71.4	53.9	69.6	70.8	76.6
18	72.4	65.4	70.6	62.2	86.5
19	72.7	72.6	70.4	63.6	84.8
20	72.0	69.8	70.4	66.5	79.8
21	69.5	50.7	64.0	65.7	82.0
22	70.5	70.6	66.7	65.3	80.0
23	68.8	73.7	66.9	61.7	77.6
24	68.5	65.7	67.3	57.4	82.0
25	67.5	67.6	67.7	58.0	77.5
26	67.2	68.2	67.2	57.3	78.0
27	66.7	67.4	67.1	57.1	76.5
28	63.2	68.3	67.8	57.1	64.4
29	63.5	67.7	68.6	57.6	64.0
30	64.2	77.7	64.5	58.6	67.9
31	60.4	77.7	64.6	58.3	55.6
32	50.8	56.0	49.7	51.1	50.7
33	50.6	56.7	49.1	50.7	51.0
34	56.6	80.8	60.6	54.9	50.9
35	57.2	82.3	56.4	54.0	57.8

English subtask detection accuracy across 4 domains

Subtask B: Multilingual Performance Across Domains

- **Dataset Breakdown**: 29 sources across 15 languages were categorized into 8 domains: News, Wikipedia, Essay, QA, Summary, Tweet, GovReport, and others
- **In-Domain Accuracy**: Structured in-domain datasets (News, Wikipedia, QA, and Summary) showed higher accuracies, with top teams achieving over 98% accuracy in QA and Wikipedia.
- **Out-of-Domain Performance**: Out-of-domain datasets (Essay, Tweet, GovReport, Other) faced greater challenges, with tweets showing the lowest performance (69.99% accuracy), reflecting difficulties in generalizing to informal text.

Rank Size	All 151,425	News 57,590	Wiki 11,687	Essay 2,201	QA 24,854	Summary 13,600	Tweet 1,325	GovR 19,736	Other 4,214
1	79.6	65.1	80.2	99.3	98.9	70.0	94.5	87.0	84.2
2	75.6	64.0	87.1	81.0	91.9	79.1	100.0	69.1	48.2
3	75.9	60.7	81.0	97.7	96.2	65.2	72.0	81.7	91.1
4	75.3	60.7	87.9	91.0	93.2	71.7	98.9	75.2	58.6
BL	74.8	61.6	85.2	97.7	94.1	58.6	94.4	76.2	83.2
5	74.7	60.2	74.7	97.7	98.9	59.7	65.3	75.0	96.2
6	74.5	59.8	79.6	90.9	95.1	82.8	95.5	62.6	82.7
7	74.4	59.8	79.7	90.7	95.2	82.1	93.8	62.9	79.4
8	73.9	58.1	81.2	98.5	92.9	73.5	29.1	81.2	70.7
9	73.5	61.1	85.0	94.7	94.5	64.8	87.8	78.7	60.3
10	73.6	60.8	77.3	94.2	95.4	61.3	91.9	80.5	86.8
11	73.3	60.2	83.9	96.7	94.9	60.0	56.0	82.4	61.8
12	73.5	62.2	81.4	93.3	95.9	64.8	41.0	83.5	68.2
13	72.0	56.3	42.3	99.2	99.2	70.9	33.7	89.0	67.3
14	71.0	56.0	55.2	97.0	92.4	76.3	0.1	81.1	85.6
15	50.3	51.0	42.4	60.0	51.2	49.7	33.9	61.9	62.1
16	71.5	59.6	44.0	97.0	99.2	59.5	57.7	89.3	58.1
17	50.2	50.8	43.2	57.7	50.7	49.9	36.6	59.8	60.8
18	69.6	55.0	45.8	97.7	92.2	71.5	2.3	82.7	85.2
19	70.5	54.5	33.5	99.1	99.1	73.1	6.4	88.7	77.6
20	70.7	60.9	41.7	93.5	99.1	63.5	45.3	86.8	61.3
21	67.9	61.7	69.9	63.6	78.1	78.0	49.4	71.8	60.7
22	67.1	57.4	51.8	83.4	94.7	61.5	100.0	80.7	20.9
23	49.7	49.1	57.0	45.5	49.1	50.3	64.5	40.1	39.4
24	52.6	45.3	35.0	83.0	72.4	67.3	99.3	46.6	17.8
25	51.0	50.4	53.0	51.0	51.8	52.0	56.1	48.4	48.9

Subtask B: Increasing Detection Difficulty with Improved Generation Prompts

- **Purpose of Improved Prompts**: The improved prompts were designed to make machine-generated text more similar to human-written text, aiming to narrow the detection gap.
- **Increased Detection Difficulty**: By using these well-designed prompts, the text became harder to distinguish, making the detection task more challenging for systems.
- **Accuracy Decline**: Detectors showed a decrease in accuracy when identifying machine-generated text with the improved prompts, with some teams experiencing up to a 15% drop in performance.

Rank Size	**All** 151,425	**Fill-gap** 32,487	**Original** 17,017	**Others** 101,921
1	79.6	91.1	94.2	73.5
2	75.6	75.9	84.0	74.1
3	75.9	89.7	92.2	68.8
4	75.3	81.5	86.9	71.4
BL	74.8	87.6	89.0	68.3
5	74.7	84.6	96.6	67.9
6	74.5	75.6	90.1	71.5
7	74.4	75.4	90.3	71.4
8	73.9	88.5	87.1	67.0
9	73.5	86.7	93.1	66.0
10	73.6	92.9	93.0	64.2
11	73.3	88.3	91.6	65.5
12	73.5	91.6	94.3	64.3
13	72.0	93.7	95.7	61.1
14	71.0	90.4	86.3	62.3
15	50.3	66.7	64.8	42.7
16	71.5	93.2	96.4	60.4
17	50.2	64.7	62.9	43.5
18	69.6	91.6	86.5	59.8
19	70.5	94.9	95.1	58.6
20	70.7	93.8	96.1	59.0
21	67.9	79.9	71.5	63.5
22	67.1	84.6	94.4	57.0
23	49.7	36.1	37.4	56.1
24	52.6	66.4	60.3	46.9
25	51.0	48.2	48.5	52.4

Subtask B: Accuracy Across Seen and Unseen Languages

- **Top-Performing Languages**: Detection accuracy is highest for seen languages, with Chinese (94.2), Russian (89.6), and Spanish (89.5) leading the results.
- **Performance on Seen Languages**: Languages like Arabic, Italian, and Dutch show slightly lower but competitive performance, demonstrating good generalization to seen languages.
- **Challenges with Unseen Languages**: Significant accuracy drops occur with unseen languages, like Hindi (51.8), due to limited exposure to linguistic patterns during training.

Rank Size	All 151,425	ZH 63,009	UR 30,505	RU 27,158	AR 10,670	IT 5,296	KK 2,471	VI 2,326	DE 1,865	NO 1,544	ID 1,200	NL 1,200	ES 1,200	HI 1,199	HE 1,182	JA 600
1	79.6	94.2	68.7	67.1	71.2	52.9	55.5	90.5	88.3	80.3	89.6	82.2	89.5	51.8	86.7	77.0
2	75.6	84.7	64.6	74.2	57.9	52.9	83.8	83.5	96.4	76.0	51.7	90.6	91.2	69.6	96.8	95.3
3	75.9	90.2	67.2	58.9	66.8	52.9	92.5	74.7	88.8	72.2	87.4	68.9	47.1	70.6	96.4	72.2
4	75.3	87.6	64.6	63.9	61.3	52.9	75.8	83.4	94.9	88.5	53.5	92.2	90.4	73.0	97.3	92.2
BL	74.8	87.3	68.4	55.3	68.4	52.9	82.8	85.3	85.2	69.8	68.2	92.5	90.5	71.3	89.3	90.0
5	74.7	90.1	64.1	56.0	69.1	52.9	62.9	87.6	59.6	69.8	93.8	81.0	90.4	69.1	96.5	95.0
6	74.5	84.2	65.0	67.9	66.8	52.9	47.5	81.8	93.5	83.2	83.9	85.9	88.9	69.1	89.8	78.2
7	74.4	84.4	64.9	67.7	65.4	52.9	47.5	82.0	92.2	85.8	83.4	85.4	89.2	68.8	90.1	75.2
8	73.9	88.3	58.7	67.0	58.4	52.9	93.0	65.9	89.6	61.6	50.5	80.7	88.0	61.4	82.7	61.2
9	73.5	85.1	67.0	59.8	60.8	52.9	90.6	87.2	82.8	78.2	48.7	78.0	83.1	54.5	89.6	74.3
10	73.6	86.0	67.6	56.0	69.1	52.9	86.8	80.4	65.0	52.8	73.8	87.4	85.4	63.5	85.7	86.0
11	73.3	87.4	63.4	58.2	55.6	52.9	89.4	79.7	87.0	66.6	73.9	82.1	87.4	70.5	93.3	79.5
12	73.5	85.3	68.0	61.5	54.3	52.9	92.7	62.0	87.8	63.7	80.3	85.3	86.3	63.0	86.2	59.5
13	72.0	93.2	55.4	63.3	55.4	52.9	93.0	65.9	5.2	25.8	71.2	50.2	50.0	61.4	1.7	61.2
14	71.0	87.0	54.3	68.7	61.2	52.8	54.7	63.8	77.1	54.7	49.7	57.1	64.9	53.5	0.0	52.0
15	50.3	50.9	52.0	49.0	53.0	50.4	52.1	49.7	33.9	33.2	49.7	50.3	50.7	50.4	32.1	50.0
16	71.5	91.3	62.4	55.5	53.7	52.9	89.4	79.7	5.3	28.9	79.9	50.2	50.0	70.3	1.9	79.5
17	50.2	50.6	51.4	49.3	52.8	50.1	52.2	50.1	35.9	34.5	49.3	50.3	50.2	50.6	34.2	53.3
18	69.6	87.4	54.5	63.8	61.1	52.9	55.7	57.0	58.2	23.1	50.3	55.2	59.3	53.7	0.0	54.3
19	70.5	92.2	51.6	65.5	56.5	52.8	54.7	63.8	4.2	23.8	70.6	50.1	50.0	53.5	0.0	52.0
20	70.7	87.6	65.6	58.3	52.0	52.9	92.7	62.0	5.0	28.2	81.7	50.2	50.0	63.0	1.9	59.5
21	67.9	71.9	51.7	80.1	55.3	78.3	48.1	63.8	93.8	82.1	72.4	83.5	84.7	52.3	31.7	63.8
22	67.1	82.5	61.5	55.3	45.8	52.9	94.2	71.6	12.0	27.9	57.5	63.3	73.6	53.5	20.3	57.2
23	49.7	49.2	48.4	50.7	47.4	49.0	50.3	49.7	65.5	63.5	50.4	51.1	49.2	51.9	64.5	52.0
24	52.6	60.7	45.7	58.9	28.8	52.9	47.5	48.1	5.8	39.8	47.7	49.5	51.2	46.0	5.8	27.0
25	51.0	51.1	49.9	51.5	50.8	50.1	50.1	52.3	55.9	54.5	52.5	54.0	49.9	52.4	53.7	52.0

Takeaways

- Most of the systems performed well on in-domain data
- Open problems:
  - **Generalization**: systems' performance drops significantly when faced with out-of-domain data and unseen languages
  - **Robustness**: systems' performance drops significantly when faced with humanized machine-generated texts
- Developing more **robust and generalizable** AI systems is a **key** for future research
- The struggle with **humanized** machine-generated texts poses a **threat** of potential misuse of LLM-based systems.

Future Events: PAN CLEF Evaluation Lab 2025, Subtask 2

"Recognition of AI text in a mixed Human-AI document"

A document written by both a human and a machine, determine which parts belong to whom

- (1) human-started, then machine-continued
- (2) mixed text, where some parts are written by a human and some are generated by a machine
- (3) human-written, then machine-polished
- (4) machine-written, then machine-polished (obfuscated) texts
- (5) human-written text

MOHAMED BIN ZAYED UNIVERSITY OF ARTIFICIAL INTELLIGENCE

Github Repo

Thank you

Mohamed bin Zayed University of Artificial Intelligence
Masdar City, Abu Dhabi, United Arab Emirates

mbzuai.ac.ae