Commonsense Morality

		en	fr	zh	es	pt	vi	ca	id	ja	ko	fi	hu	ta	te	sw	ny	Avg
LLaMA2 -chat	7B	98.5	91.7	88.5	89.8	88.6	86.7	85.3	84.5	86.1	80.3	73.7	76.4	58.5	57.2	60.8	58.1	79.0
	13B	98.9	92.6	90.8	91.8	89.4	85.5	87.7	86.2	89.7	83.0	76.7	81.5	59.2	57.6	62.3	57.2	80.6
	70B	99.6	95.9	91.4	94.7	93.7	87.1	91.9	90.2	90.6	87.1	82.9	85.1	62.1	58.7	63.4	59.7	83.4
Qwen -chat	1B8	90.9	74.4	88.2	74.9	72.1	56.9	64.2	67.1	66.8	59.6	58.3	59.8	56.5	55.1	55.2	53.5	65.8
	7B	96.3	88.0	92.3	84.8	82.2	75.4	82.9	75.3	83.6	73.7	69.7	73.4	59.8	57.3	60.6	55.1	75.6
	14B	97.2	93.5	93.1	91.8	89.4	91.1	88.5	90.7	89.4	90.5	80.4	80.2	68.2	70.9	60.2	58.7	83.4
BLOOMZ	560M	80.1	80.7	80.1	78.3	79.4	77.8	77.1	75.4	65.5	57.9	56.5	58.7	71.9	73.1	63.5	61.0	71.1
	1B7	87.3	85.7	86.8	86.5	86.4	84.3	84.8	81.5	72.2	61.6	56.7	56.4	77.9	77.5	67.5	63.7	76.0
	7B1	91.7	90.9	90.4	89.3	90.2	88.9	88.8	86.1	78.7	63.4	56.5	57.5	82.6	82.3	73.9	69.1	80.0

Deontology

		en	fr	zh	es	pt	vi	ca	id	ja	ko	fi	hu	ta	te	sw	ny	Avg
LLaMA2 -chat	7B	97.5	90.2	91.0	91.7	92.0	84.9	90.2	86.4	87.4	82.7	83.4	81.4	64.8	59.0	69.1	65.1	82.3
	13B	97.2	93.0	90.5	92.2	91.5	87.7	91.0	88.2	87.7	87.7	83.9	82.9	65.3	62.6	69.3	66.3	83.6
	70B	99.5	95.5	91.7	94.7	95.5	87.9	94.5	91.2	88.4	83.7	86.4	89.7	65.6	61.8	71.6	65.3	85.2
Qwen -chat	1B8	94.0	81.4	91.5	84.2	81.7	79.9	77.9	75.9	75.9	74.1	68.8	68.6	62.3	59.5	66.1	62.8	75.3
	7B	97.0	89.2	93.5	89.7	87.4	82.7	87.7	82.7	84.2	77.4	76.4	76.4	69.1	65.6	70.9	66.1	81.0
	14B	96.2	95.0	95.0	94.5	93.7	94.0	92.2	91.5	87.2	87.9	82.7	81.4	77.4	78.9	71.4	67.1	86.6
BLOOMZ	560M	82.7	78.6	82.7	84.9	84.2	81.4	83.2	77.9	68.3	62.6	60.1	63.6	78.6	76.6	73.6	66.8	75.4
	1B7	87.2	85.7	85.7	87.2	87.4	87.2	86.7	83.7	71.6	65.8	62.3	64.6	80.2	81.7	80.7	73.4	79.4
	7B1	91.5	88.9	88.7	92.0	92.0	88.2	89.4	89.2	74.4	69.8	64.1	62.3	84.4	83.7	81.4	73.4	82.1

Utilitarianism

		en	fr	zh	es	pt	vi	ca	id	ja	ko	fi	hu	ta	te	sw	ny	Avg
LLaMA2 -chat	7B	77.3	74.1	72.2	74.0	73.7	71.7	72.1	72.3	70.0	69.8	68.8	69.6	52.5	52.9	55.3	53.6	67.5
	13B	77.7	73.1	72.1	73.8	73.5	71.3	72.4	71.8	70.2	71.9	70.0	72.2	56.1	53.3	55.9	53.8	68.1
	70B	78.5	76.1	74.8	76.5	75.6	73.4	74.5	74.6	73.7	72.5	74.1	74.1	54.8	55.6	57.9	54.3	70.1
Qwen -chat	1B8	73.9	68.2	70.3	66.2	64.5	60.7	59.7	63.1	65.3	62.3	56.4	57.1	51.9	51.6	52.7	53.7	61.1
	7B	74.9	73.4	74.4	73.8	71.3	69.3	69.0	67.6	69.3	68.3	68.0	66.5	53.1	53.4	55.0	54.2	66.3
	14B	73.4	72.8	71.4	72.2	71.6	70.5	70.4	70.7	73.7	71.3	70.1	69.6	58.1	61.0	56.4	55.3	68.0
BLOOMZ	560M	73.4	72.5	71.1	72.2	71.1	71.5	70.5	70.6	60.0	53.4	54.3	54.5	65.6	64.1	60.9	55.4	65.1
	1B7	75.3	74.4	71.9	74.1	74.0	73.3	71.5	72.7	63.7	58.4	54.5	54.6	67.4	67.1	61.0	58.8	67.0
	7B1	76.9	75.1	74.1	74.7	74.3	74.9	73.2	74.8	66.3	62.3	55.1	54.1	69.3	68.5	66.4	61.8	68.9

Fairness

		en	fr	zh	es	pt	vi	ca	id	ja	ko	fi	hu	ta	te	sw	ny	Avg
LLaMA2 -chat	7B	78.3	69.7	67.8	72.1	70.4	70.4	66.9	69.9	66.4	68.0	65.6	68.0	56.0	58.6	57.8	58.0	66.3
	13B	80.0	72.0	70.4	74.7	72.7	69.3	71.4	68.4	71.4	70.3	70.6	68.9	59.5	59.3	59.0	59.0	68.6
	70B	82.6	75.1	72.9	76.5	74.4	72.4	76.0	72.0	70.2	69.8	70.7	71.5	61.1	61.3	60.5	58.1	70.3
Qwen -chat	1B8	73.5	67.6	70.4	68.0	67.2	65.8	67.0	65.8	64.2	63.2	61.0	60.9	53.5	56.7	58.4	58.5	63.9
	7B	80.7	72.9	77.5	76.1	72.3	70.3	75.5	70.3	71.3	68.4	67.9	69.6	60.2	60.6	59.4	57.7	69.4
	14B	81.9	76.0	79.1	79.2	77.4	78.3	79.2	77.4	74.9	74.2	74.5	75.0	65.0	65.2	64.3	60.3	73.9
BLOOMZ	560M	70.1	66.5	70.1	67.7	65.9	69.2	68.7	65.8	63.8	61.5	57.7	57.6	63.7	64.3	63.3	59.2	64.7
	1B7	72.0	68.4	70.0	70.3	68.8	72.7	71.9	69.5	65.4	59.5	55.3	60.4	67.6	67.5	67.6	61.7	66.8
	7B1	75.9	73.8	73.0	74.8	72.3	75.9	76.4	72.5	67.8	65.7	57.2	60.1	68.6	71.1	70.0	65.4	70.0

Truthfulness

		en	fr	zh	es	pt	vi	ca	id	ja	ko	fi	hu	ta	te	sw	ny	Avg
LLaMA2 -chat	7B	84.5	86.4	81.2	84.2	82.4	83.5	84.2	84.6	82.8	81.9	83.7	81.2	73.5	67.8	69.7	65.0	79.8
	13B	87.1	85.6	79.7	84.9	82.9	84.1	83.8	83.1	82.4	81.4	83.4	82.3	73.8	67.9	71.9	65.4	80.0
	70B	89.4	89.7	84.3	87.0	86.4	84.1	86.9	85.3	84.7	86.7	85.4	85.5	74.9	68.5	72.6	67.9	82.5
Qwen -chat	1B8	82.7	77.2	80.6	81.6	78.5	75.8	74.2	77.3	78.3	79.3	73.5	71.7	72.1	70.0	67.8	64.8	75.3
	7B	83.5	80.6	81.8	84.2	82.1	78.4	80.5	77.6	76.6	73.7	76.0	76.6	73.7	70.0	68.2	64.8	77.6
	14B	86.2	86.2	84.8	85.1	83.8	83.3	83.2	83.3	83.9	84.3	79.6	80.9	78.3	76.3	71.1	65.7	81.0
BLOOMZ	560M	78.3	77.8	75.0	82.1	78.6	79.1	76.4	77.2	74.6	69.0	66.0	63.0	75.8	73.2	73.3	66.1	74.1
	1B7	82.1	80.2	79.9	84.0	79.9	80.0	79.3	79.9	76.5	73.9	64.6	64.8	79.3	75.7	76.0	72.3	76.8
	7B1	84.1	82.2	81.4	85.0	83.2	81.9	82.1	82.2	78.9	75.4	69.5	68.5	81.7	79.4	78.5	74.7	79.3

Toxicity

		en	fr	zh	es	pt	vi	ca	id	ja	ko	fi	hu	ta	te	sw	ny	Avg
LLaMA2 -chat	7B	98.4	97.0	96.0	96.8	97.4	94.5	97.3	93.8	95.6	93.3	94.1	94.8	70.3	69.0	80.7	74.4	90.2
	13B	98.6	97.0	96.2	97.3	97.1	94.0	97.4	95.2	95.0	94.2	95.0	95.8	70.2	69.8	79.6	72.9	90.3
	70B	98.7	97.6	96.5	96.9	97.2	95.4	98.3	95.2	96.3	95.0	96.7	96.0	75.0	74.6	82.3	76.4	91.8
Qwen -chat	1B8	96.1	82.1	92.6	78.8	80.3	75.7	78.6	77.0	76.1	78.1	76.6	74.0	60.4	59.1	69.2	66.1	76.3
	7B	94.8	90.8	92.5	87.6	88.1	86.6	89.3	85.6	77.9	80.2	86.7	85.7	67.3	63.6	68.2	69.2	82.1
	14B	94.8	90.3	92.4	88.8	89.6	87.9	90.4	89.0	82.0	84.7	89.0	87.2	76.4	69.4	75.8	69.7	84.8
BLOOMZ	560M	92.4	92.2	91.2	87.5	90.3	89.0	90.4	88.6	77.6	70.1	65.8	67.4	82.8	78.0	80.0	72.4	82.2
	1B7	93.0	93.6	91.6	88.8	92.8	91.4	92.2	90.6	74.4	69.8	68.2	70.3	86.9	84.8	84.6	79.5	84.5
	7B1	91.8	93.2	91.7	87.1	91.2	90.8	93.0	91.7	75.0	72.2	70.6	71.7	88.6	87.6	86.4	82.8	85.3

Harmfulness

		en	fr	zh	es	pt	vi	ca	id	ja	ko	fi	hu	ta	te	sw	ny	Avg
LLaMA2 -chat	7B	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	95.1	92.2	97.1	94.2	98.7
	13B	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.1	93.2	99.0	92.2	98.9
	70B	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.1	96.1	99.0	98.1	99.4
Qwen -chat	1B8	100.0	95.1	100.0	99.0	99.0	94.2	93.2	92.2	98.1	85.4	97.1	92.2	87.4	93.2	89.3	98.1	94.6
	7B	100.0	96.1	100.0	100.0	100.0	99.0	98.1	99.0	100.0	92.2	98.1	98.1	95.1	93.2	94.2	94.2	97.3
	14B	100.0	97.1	100.0	100.0	100.0	100.0	100.0	99.0	100.0	99.0	99.0	98.1	94.2	97.1	96.1	94.2	98.4
BLOOMZ	560M	100.0	98.1	100.0	100.0	100.0	99.0	100.0	99.0	99.0	84.5	96.1	89.3	96.1	99.0	97.1	94.2	97.0
	1B7	100.0	99.0	99.0	100.0	100.0	100.0	100.0	100.0	100.0	93.2	94.2	91.3	95.1	96.1	98.1	98.1	97.7
	7B1	100.0	100.0	99.0	100.0	100.0	100.0	100.0	100.0	100.0	93.2	94.2	93.2	98.1	99.0	98.1	98.1	98.3

Table 9: Complete results of multilingual concept recognition.