Llava-1.5

easy

Random(18:20)

TP FP TN FN

352 118 98 32

Accuracy: 0.75

Precision: 0.7489361702127659

Recall: 0.9166666666666666

F1 score: 0.8243559718969554

Yes ratio: 0.7833333333333333

Popular(18:30)

TP FP TN FN

352 30 186 32

Accuracy: 0.8966666666666666

Precision: 0.9214659685863874

Recall: 0.9166666666666666

F1 score: 0.9190600522193211

Yes ratio: 0.6366666666666667

Adversarial(17:54)

TP FP TN FN

352 95 121 32

Accuracy: 0.7883333333333333

Precision: 0.7874720357941835

Recall: 0.9166666666666666

F1 score: 0.8471720818291215

Yes ratio: 0.745

Medium

Random(24:21)

TP FP TN FN

346 196 211 47

Accuracy: 0.69625

Precision: 0.6383763837638377

Recall: 0.8804071246819338

F1 score: 0.7401069518716579

Yes ratio: 0.6775

Popular(24:30)

TP FP TN FN

346 34 373 47

Accuracy: 0.89875

Precision: 0.9105263157894737

Recall: 0.8804071246819338

F1 score: 0.8952134540750323

Yes ratio: 0.475

Adversarial(23:57)

TP FP TN FN

346 148 259 47

Accuracy: 0.75625

Precision: 0.7004048582995951

Recall: 0.8804071246819338

F1 score: 0.7801578354002254

Yes ratio: 0.6175

Hard

Random(31:14)

TP FP TN FN

359 240 352 49

Accuracy: 0.711

Precision: 0.5993322203672788

Recall: 0.8799019607843137

F1 score: 0.7130089374379345

Yes ratio: 0.599

Popular(30:47)

TP FP TN FN

359 176 416 49

Accuracy: 0.775

Precision: 0.6710280373831776

Recall: 0.8799019607843137

F1 score: 0.7613997879109226

Yes ratio: 0.535

Adversarial(30:48)

TP FP TN FN

359 157 435 49

Accuracy: 0.794

Precision: 0.6957364341085271

Recall: 0.8799019607843137

F1 score: 0.777056277056277

Yes ratio: 0.516

instructblip

Easy

Random(06:48)

TP FP TN FN

283 51 165 101

Accuracy: 0.7466666666666667

Precision: 0.8473053892215568

Recall: 0.7369791666666666

F1 score: 0.7883008356545961

Yes ratio: 0.5566666666666666

Popular(06:53)

TP FP TN FN

283 49 167 101

Accuracy: 0.75

Precision: 0.8524096385542169

Recall: 0.7369791666666666

F1 score: 0.7905027932960894

Yes ratio: 0.5533333333333333

Adversarial(06:44)

TP FP TN FN

283 94 122 101

Accuracy: 0.675

Precision: 0.7506631299734748

Recall: 0.7369791666666666

F1 score: 0.7437582128777923

Yes ratio: 0.6283333333333333

Medium

Random(08:53)

TP FP TN FN

286 109 298 107

Accuracy: 0.73

Precision: 0.7240506329113924

Recall: 0.727735368956743

F1 score: 0.7258883248730964

Yes ratio: 0.49375

Popular(09:13)

TP FP TN FN

286 42 365 107

Accuracy: 0.81375

Precision: 0.8719512195121951

Recall: 0.727735368956743

F1 score: 0.7933425797503467

Yes ratio: 0.41

Adversarial(09:00)

TP FP TN FN

286 120 287 107

Accuracy: 0.71625

Precision: 0.7044334975369458

Recall: 0.727735368956743

F1 score: 0.7158948685857323

Yes ratio: 0.5075

Hard

Random(10:57)

TP FP TN FN

292 101 491 116

Accuracy: 0.783

Precision: 0.7430025445292621

Recall: 0.7156862745098039

F1 score: 0.7290886392009989

Yes ratio: 0.393

Popular(10:53)

TP FP TN FN

292 86 506 116

Accuracy: 0.798

Precision: 0.7724867724867724

Recall: 0.7156862745098039

F1 score: 0.7430025445292621

Yes ratio: 0.378

Adversarial(11:03)

TP FP TN FN

292 108 484 116

Accuracy: 0.776

Precision: 0.73

Recall: 0.7156862745098039

F1 score: 0.7227722772277227

Yes ratio: 0.4

Minigpt4(vicuna-7b)

easy

Random(06:30)

TP FP TN FN

378 185 31 6

Accuracy: 0.6816666666666666

Precision: 0.6714031971580817

Recall: 0.984375

F1 score: 0.7983104540654699

Yes ratio: 0.9383333333333334

Popular(06:36)

TP FP TN FN

374 181 35 10

Accuracy: 0.6816666666666666

Precision: 0.6738738738738739

Recall: 0.9739583333333334

F1 score: 0.7965921192758254

Yes ratio: 0.925

Adversarial(06:33)

TP FP TN FN

374 191 25 10

Accuracy: 0.665

Precision: 0.6619469026548672

Recall: 0.9739583333333334

F1 score: 0.7881981032665963

Yes ratio: 0.9416666666666667

Medium

Random(08:40)

TP FP TN FN

370 381 26 23

Accuracy: 0.495

Precision: 0.49267643142476697

Recall: 0.9414758269720102

F1 score: 0.6468531468531468

Yes ratio: 0.93875

Popular(08:39)

TP FP TN FN

370 338 69 23

Accuracy: 0.54875

Precision: 0.5225988700564972

Recall: 0.9414758269720102

F1 score: 0.6721162579473206

Yes ratio: 0.885

Adversarial(08:40)

TP FP TN FN

370 360 47 23

Accuracy: 0.52125

Precision: 0.5068493150684932

Recall: 0.9414758269720102

F1 score: 0.6589492430988423

Yes ratio: 0.9125

Hard

Random(10:49)

TP FP TN FN

361 542 50 47

Accuracy: 0.411

Precision: 0.3997785160575858

Recall: 0.8848039215686274

F1 score: 0.5507246376811594

Yes ratio: 0.903

Popular(10:51)

TP FP TN FN

361 507 85 47

Accuracy: 0.446

Precision: 0.41589861751152074

Recall: 0.8848039215686274

F1 score: 0.5658307210031348

Yes ratio: 0.868

Adversarial(10:49)

TP FP TN FN

361 506 86 47

Accuracy: 0.447

Precision: 0.41637831603229525

Recall: 0.8848039215686274

F1 score: 0.5662745098039215

Yes ratio: 0.867

Shikra()

Easy

Random(21:34)

TP FP TN FN

339 129 87 45

Accuracy: 0.71

Precision: 0.7243589743589743

Recall: 0.8828125

F1 score: 0.7957746478873239

Yes ratio: 0.78

Popular(21:26)

TP FP TN FN

339 62 154 45

Accuracy: 0.8216666666666667

Precision: 0.8453865336658354

Recall: 0.8828125

F1 score: 0.8636942675159235

Yes ratio: 0.6683333333333333

Adversarial(21:26)

TP FP TN FN

339 100 116 45

Accuracy: 0.7583333333333333

Precision: 0.7722095671981777

Recall: 0.8828125

F1 score: 0.8238153098420413

Yes ratio: 0.7316666666666667

Medium

Random(28:41)

TP FP TN FN

339 217 190 54

Accuracy: 0.66125

Precision: 0.6097122302158273

Recall: 0.8625954198473282

F1 score: 0.7144362486828241

Yes ratio: 0.695

Popular(28:39)

TP FP TN FN

339 97 310 54

Accuracy: 0.81125

Precision: 0.7775229357798165

Recall: 0.8625954198473282

F1 score: 0.8178528347406513

Yes ratio: 0.545

Adversarial(28:40)

TP FP TN FN

339 161 246 54

Accuracy: 0.73125

Precision: 0.678

Recall: 0.8625954198473282

F1 score: 0.7592385218365062

Yes ratio: 0.625

Hard

Random(35:30)

TP FP TN FN

345 256 336 63

Accuracy: 0.681

Precision: 0.5740432612312812

Recall: 0.8455882352941176

F1 score: 0.6838453914767096

Yes ratio: 0.601

Popular(35:53)

TP FP TN FN

345 174 418 63

Accuracy: 0.763

Precision: 0.6647398843930635

Recall: 0.8455882352941176

F1 score: 0.744336569579288

Yes ratio: 0.519

Adversarial(35:52)

TP FP TN FN

345 182 410 63

Accuracy: 0.755

Precision: 0.6546489563567363

Recall: 0.8455882352941176

F1 score: 0.7379679144385026

Yes ratio: 0.527

Minigpt4(llama2-7b)

Easy

Random(06:48)

TP FP TN FN

360 183 33 24

Accuracy: 0.655

Precision: 0.6629834254143646

Recall: 0.9375

F1 score: 0.7766990291262136

Yes ratio: 0.905

Popular(06:48)

TP FP TN FN

360 147 69 24

Accuracy: 0.715

Precision: 0.7100591715976331

Recall: 0.9375

F1 score: 0.808080808080808

Yes ratio: 0.845

Adversarial(06:47)

TP FP TN FN

360 171 45 24

Accuracy: 0.675

Precision: 0.6779661016949152

Recall: 0.9375

F1 score: 0.7868852459016393

Yes ratio: 0.885

Meidum

Random(09:03)

TP FP TN FN

355 366 41 38

Accuracy: 0.495

Precision: 0.492371705963939

Recall: 0.9033078880407125

F1 score: 0.6373429084380611

Yes ratio: 0.90125

Popular(09:05)

TP FP TN FN

355 288 119 38

Accuracy: 0.5925

Precision: 0.552099533437014

Recall: 0.9033078880407125

F1 score: 0.6853281853281853

Yes ratio: 0.80375

Adversarial(09:03)

TP FP TN FN

355 316 91 38

Accuracy: 0.5575

Precision: 0.5290611028315947

Recall: 0.9033078880407125

F1 score: 0.6672932330827068

Yes ratio: 0.83875

Hard

Random(11:21)

TP FP TN FN

366 522 70 42

Accuracy: 0.436

Precision: 0.41216216216216217

Recall: 0.8970588235294118

F1 score: 0.5648148148148148

Yes ratio: 0.888

Popular(11:15)

TP FP TN FN

366 451 141 42

Accuracy: 0.507

Precision: 0.4479804161566707

Recall: 0.8970588235294118

F1 score: 0.5975510204081633

Yes ratio: 0.817

Adversarial(11:19)

TP FP TN FN

366 454 138 42

Accuracy: 0.504

Precision: 0.44634146341463415

Recall: 0.8970588235294118

F1 score: 0.5960912052117263

Yes ratio: 0.82

Instructblip(13b)

Easy

Random(08:44)

TP FP TN FN

344 145 71 40

Accuracy: 0.6916666666666667

Precision: 0.7034764826175869

Recall: 0.8958333333333334

F1 score: 0.7880870561282932

Yes ratio: 0.815

Popular(08:43)

TP FP TN FN

344 57 159 40

Accuracy: 0.8383333333333334

Precision: 0.85785536159601

Recall: 0.8958333333333334

F1 score: 0.8764331210191082

Yes ratio: 0.6683333333333333

Adversarial(08:29)

TP FP TN FN

344 128 88 40

Accuracy: 0.72

Precision: 0.7288135593220338

Recall: 0.8958333333333334

F1 score: 0.8037383177570093

Yes ratio: 0.7866666666666666

Medium

Random(11:52)

TP FP TN FN

332 254 153 61

Accuracy: 0.60625

Precision: 0.5665529010238908

Recall: 0.8447837150127226

F1 score: 0.6782431052093973

Yes ratio: 0.7325

Popular(11:48)

TP FP TN FN

332 120 287 61

Accuracy: 0.77375

Precision: 0.7345132743362832

Recall: 0.8447837150127226

F1 score: 0.7857988165680474

Yes ratio: 0.565

Adversarial(11:47)

TP FP TN FN

332 207 200 61

Accuracy: 0.665

Precision: 0.6159554730983302

Recall: 0.8447837150127226

F1 score: 0.7124463519313305

Yes ratio: 0.67375

Hard

Random(15:53)

TP FP TN FN

319 311 281 89

Accuracy: 0.6

Precision: 0.5063492063492063

Recall: 0.7818627450980392

F1 score: 0.6146435452793835

Yes ratio: 0.63

Popular(14:45)

TP FP TN FN

319 236 356 89

Accuracy: 0.675

Precision: 0.5747747747747748

Recall: 0.7818627450980392

F1 score: 0.6625129802699896

Yes ratio: 0.555

Adversarial(15:03)

TP FP TN FN

319 256 336 89

Accuracy: 0.655

Precision: 0.5547826086956522

Recall: 0.7818627450980392

F1 score: 0.6490335707019328

Yes ratio: 0.575

Llava-1.5(13b)

Easy

Random(24:33)

TP FP TN FN

368 138 78 16

Accuracy: 0.7433333333333333

Precision: 0.7272727272727273

Recall: 0.9583333333333334

F1 score: 0.8269662921348315

Yes ratio: 0.8433333333333334

Popular(24:33)

TP FP TN FN

368 76 140 16

Accuracy: 0.8466666666666667

Precision: 0.8288288288288288

Recall: 0.9583333333333334

F1 score: 0.8888888888888888

Yes ratio: 0.74

Adversarial(24:30)

TP FP TN FN

368 115 101 16

Accuracy: 0.7816666666666666

Precision: 0.7619047619047619

Recall: 0.9583333333333334

F1 score: 0.8489042675893886

Yes ratio: 0.805

Medium

Random(32:43)

TP FP TN FN

363 238 169 30

Accuracy: 0.665

Precision: 0.6039933444259568

Recall: 0.9236641221374046

F1 score: 0.7303822937625755

Yes ratio: 0.75125

Popular(32:40)

TP FP TN FN

363 92 315 30

Accuracy: 0.8475

Precision: 0.7978021978021979

Recall: 0.9236641221374046

F1 score: 0.8561320754716982

Yes ratio: 0.56875

Adversarial(32:40)

TP FP TN FN

363 194 213 30

Accuracy: 0.72

Precision: 0.651705565529623

Recall: 0.9236641221374046

F1 score: 0.7642105263157895

Yes ratio: 0.69625

Hard

Random(40:51)

TP FP TN FN

359 287 305 49

Accuracy: 0.664

Precision: 0.5557275541795665

Recall: 0.8799019607843137

F1 score: 0.6812144212523719

Yes ratio: 0.646

Popular(40:49)

TP FP TN FN

359 226 366 49

Accuracy: 0.725

Precision: 0.6136752136752137

Recall: 0.8799019607843137

F1 score: 0.7230614300100705

Yes ratio: 0.585

Adversarial(40:49)

TP FP TN FN

359 200 392 49

Accuracy: 0.751

Precision: 0.6422182468694096

Recall: 0.8799019607843137

F1 score: 0.7425025853154085

Yes ratio: 0.559

Minigpt4(vicuna-13b)

Easy

Random(11:08)

TP FP TN FN

230 69 147 154

Accuracy: 0.6283333333333333

Precision: 0.7692307692307693

Recall: 0.5989583333333334

F1 score: 0.6734992679355783

Yes ratio: 0.49833333333333335

Popular(11:09)

TP FP TN FN

230 52 164 154

Accuracy: 0.6566666666666666

Precision: 0.8156028368794326

Recall: 0.5989583333333334

F1 score: 0.6906906906906906

Yes ratio: 0.47

Adversarial(11:09)

TP FP TN FN

230 53 163 154

Accuracy: 0.655

Precision: 0.8127208480565371

Recall: 0.5989583333333334

F1 score: 0.689655172413793

Yes ratio: 0.4716666666666667

Medium

Random(14:50)

TP FP TN FN

222 137 270 171

Accuracy: 0.615

Precision: 0.6183844011142061

Recall: 0.5648854961832062

F1 score: 0.5904255319148936

Yes ratio: 0.44875

Popular(14:52)

TP FP TN FN

222 94 313 171

Accuracy: 0.66875

Precision: 0.7025316455696202

Recall: 0.5648854961832062

F1 score: 0.6262341325811002

Yes ratio: 0.395

Adversarial(14:52)

TP FP TN FN

222 135 272 171

Accuracy: 0.6175

Precision: 0.6218487394957983

Recall: 0.5648854961832062

F1 score: 0.5920000000000001

Yes ratio: 0.44625

Hard

Random(18:30)

TP FP TN FN

183 156 436 225

Accuracy: 0.619

Precision: 0.5398230088495575

Recall: 0.4485294117647059

F1 score: 0.4899598393574297

Yes ratio: 0.339

Popular(18:38)

TP FP TN FN

183 126 466 225

Accuracy: 0.649

Precision: 0.5922330097087378

Recall: 0.4485294117647059

F1 score: 0.5104602510460251

Yes ratio: 0.309

Adversarial(18:38)

TP FP TN FN

183 137 455 225

Accuracy: 0.638

Precision: 0.571875

Recall: 0.4485294117647059

F1 score: 0.5027472527472527

Yes ratio: 0.32