

Appendix

A Survey Results

In the second section of our three-part investigation, we conducted a user study to understand developers’ experiences with model requirement changes and their actions. In the paper, we observed the relation between developer experience and the probability of facing model changes. Here, we present the combined participant distribution for all changes in Figure 1.

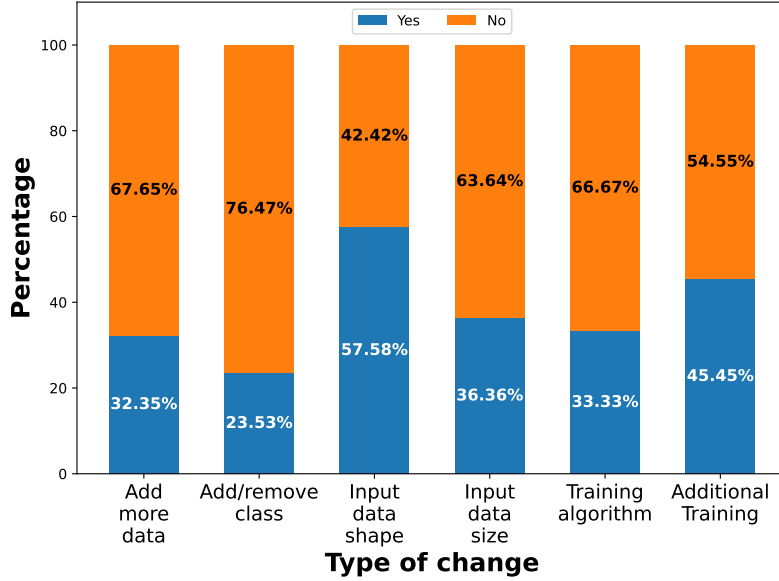


Figure 1: Overall participant distribution for different types of change

Next, we present the results for Q10 and Q12 of our survey questionnaire, as depicted in Table II in the paper. Figure 2 shows the distribution of participants when asked about the frequency with which they faced the need to retrain the models. Figure 3 illustrates the answer to Q12. It shows that around 84% of participants were interested in learning about methods to avoid retraining.

B Benchmarking Results

Here, we present the results for Section IV.C. We perform each experiment twice.

B.1 Add more data to Retrain

This experiment is performed for output-based decomposition. We first train a model and report the monolithic model accuracy as **MMA Before**. We then decompose it into modules and report the composed accuracy (**CMA Before**). Lastly, we apply the retraining on the concerned modules and report the composed accuracy **CMA After**. The results are summarized in Table 1. Here, we only provide results for dense model with 4 layers. We also performed experiments on models with 1, 2 and 3 hidden layers. The complete results are in our replication package [1].

B.2 Add/ Remove a Class

Neural module networks can restrict changes to some modules when classes are removed. Output-based decomposition can do so when classes are removed and added. Here, we present the results

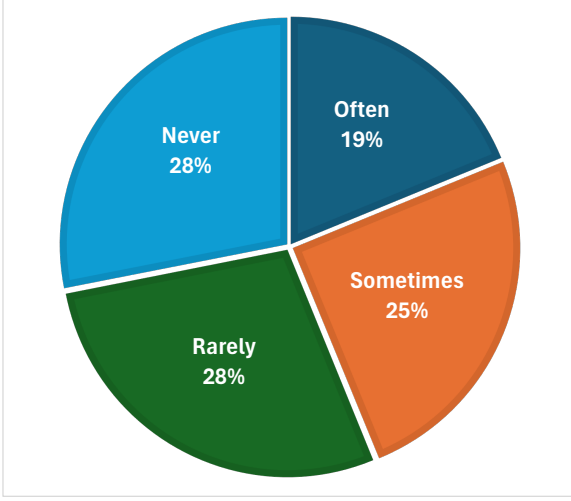


Figure 2: Frequency of retraining requirement

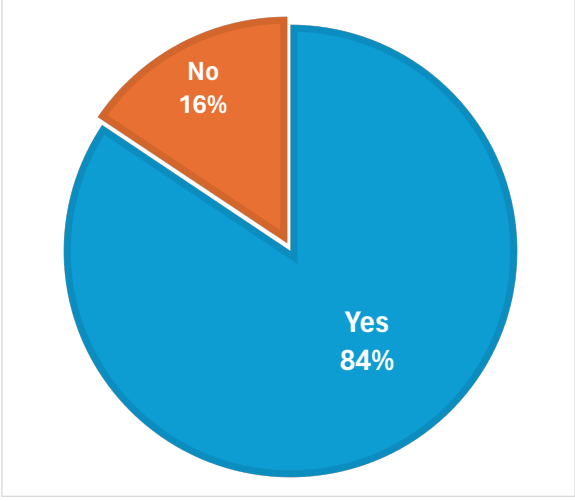


Figure 3: Interest in decomposition styles

of this type of change. Table 2 and Table 3 summarize the results for neural module networks and output-based decomposition, respectively. For output-based decomposition, the results for models with 1, 2, and 3 layers are present in our replication package.

B.3 Change in Data Shape

Here, we present the evaluation results for layer-based decomposition since it is the most effective in this change type. The results are presented in Table 4. We use the following terminologies in the table: **Pre-MA**: Original model accuracy before the change, **OMA**: Modified model accuracy on the changed data, **Po-MA**: Accuracy of the newly trained model with the changed data, **Po-TP**: # of trainable parameters in the newly trained model, **MNTP**: # of non-trainable parameters in the modified model, **MTP**: # of trainable parameters in the modified model.

B.4 Change in Data Type

Table 5 presents the results for layer-based decomposition. We use the following terminologies: **CIDT**: Changed Input Data Type, **MIDT**: Model Trained Data Type, **Pre-MA**: Model Accuracy Before Change, **Po-MA**: Model Accuracy After Change, **OMA**: Modified Model Accuracy, **S<:T** - S is a subtype of type T.

Table 1: Add more data to retrain the monolithic model vs. retraining the affected module

Class	Accuracy	MNIST-4	EMNIST-4	FMNIST-4	KMNIST-4
0	MMA Before	94.52%	84.81%	81.47%	80.60%
	CMA Before	94.55%	84.85%	81.20%	80.65%
	MMA After	95.88%	92.54%	87.20%	86.62%
	CMA After	96.41%	91.28%	85.84%	85.05%
1	MMA Before	95.28%	87.25%	85.54%	79.23%
	CMA Before	95.29%	87.34%	85.59%	79.13%
	MMA After	97.11%	92.64%	87.60%	85.80%
	CMA After	96.84%	92.09%	87.47%	84.87%
2	MMA Before	92.89%	84.36%	81.84%	80.21%
	CMA Before	92.94%	84.48%	81.86%	80.27%
	MMA After	96.58%	92.78%	87.70%	86.43%
	CMA After	96.26%	88.93%	85.51%	85.76%
3	MMA Before	92.18%	86.73%	81.71%	79.77%
	CMA Before	92.21%	86.81%	81.60%	79.88%
	MMA After	96.38%	92.71%	87.43%	86.16%
	CMA After	96.29%	92.36%	87.60%	84.67%
4	MMA Before	91.41%	87.40%	80.95%	78.64%
	CMA Before	91.41%	87.05%	80.84%	78.62%
	MMA After	96.69%	92.11%	86.92%	84.84%
	CMA After	96.13%	90.38%	83.76%	83.98%
5	MMA Before	89.75%	87.64%	83.28%	79.12%
	CMA Before	89.69%	87.64%	83.19%	79.02%
	MMA After	96.84%	91.79%	87.88%	84.66%
	CMA After	95.15%	91.86%	87.61%	84.38%
6	MMA Before	93.86%	85.36%	82.99%	79.51%
	CMA Before	93.84%	85.38%	82.96%	79.53%
	MMA After	96.63%	92.51%	87.55%	85.70%
	CMA After	96.47%	91.15%	86.22%	83.52%
7	MMA Before	89.52%	87.84%	83.90%	80.80%
	CMA Before	89.52%	87.65%	83.81%	80.75%
	MMA After	96.61%	92.69%	86.95%	85.59%
	CMA After	94.69%	92.24%	85.07%	84.60%
8	MMA Before	92.81%	89.11%	83.55%	79.20%
	CMA Before	92.88%	89.14%	83.52%	79.14%
	MMA After	96.03%	92.19%	87.79%	85.01%
	CMA After	96.67%	91.95%	87.15%	83.92%
9	MMA Before	91.21%	85.33%	83.32%	80.44%
	CMA Before	91.22%	85.40%	83.10%	80.42%
	MMA After	96.96%	92.48%	87.11%	84.91%
	CMA After	95.57%	92.41%	83.84%	85.59%

MMA: {Before/After}: Monolithic Model Accuracy {before/after} the Change, CMA: {Before/After}: Composed Model Accuracy {before/after} the Change.

Table 2: Adding and removing classes using NMNs.

Dataset	VQA-MNIST-1				VQA-MNIST-2	
Changes	Add Class		Remove Class		Add Class	Remove Class
Tasks	AAA	GAA	AAA	GAA	AAA	AAA
0	74.19%	80.32%	74.50%	80.22%	76.11%	74.34%
1	71.69%	78.75%	70.69%	78.72%	78.04%	74.37%
2	72.73%	79.22%	73.22%	79.19%	76.64%	74.75%
3	72.64%	80.78%	71.95%	79.85%	76.42%	76.20%
4	71.70%	79.56%	71.68%	79.74%	77.73%	75.29%
5	72.63%	80.77%	72.95%	80.88%	76.84%	75.27%
6	71.68%	80.04%	71.92%	81.66%	76.55%	74.90%
7	70.17%	78.62%	70.57%	80.01%	76.75%	74.69%
8	74.12%	80.37%	73.75%	80.97%	77.58%	75.00%
9	71.92%	82.44%	71.62%	83.27%	78.04%	75.13%
OM	75.45%	80.32%	75.45%	80.32%	74.20%	74.20%

OM: Original Model. AAA: All-all-all, GAA: Group-all-all

Table 3: Adding/Removing class for only the affected class vs. monolithic model

Class	Accuracy	MNIST-4		EMNIST-4		FMNIST-4		KMNIST-4	
		Add Class	Remove Class	Add Class	Remove Class	Add Class	Remove Class	Add Class	Remove Class
0	MMA Before	96.90%	96.9%	93.49%	92.3%	90.07%	96.7%	85.19%	86.1%
	CMA Before	96.90%	96.9%	93.49%	92.3%	90.07%	96.7%	85.19%	86.1%
	MMA After	96.53%	96.8%	92.01%	93.4%	87.20%	96.2%	84.71%	85.9%
	CMA After	96.87%	97.0%	91.86%	93.4%	87.20%	96.6%	84.96%	86.1%
1	MMA Before	96.39%	96.6%	93.18%	92.7%	86.94%	95.5%	85.98%	85.6%
	CMA Before	96.39%	96.6%	93.18%	92.7%	86.94%	95.5%	85.98%	85.6%
	MMA After	96.36%	96.8%	92.50%	92.7%	87.36%	96.8%	85.68%	84.8%
	CMA After	96.48%	96.4%	91.54%	93.5%	87.65%	95.3%	84.69%	86.9%
2	MMA Before	96.88%	96.5%	93.21%	92.6%	90.12%	96.8%	87.36%	86.2%
	CMA Before	96.88%	96.5%	93.21%	92.6%	90.12%	96.8%	87.36%	86.2%
	MMA After	96.78%	97.0%	92.64%	92.8%	87.65%	96.7%	83.99%	87.3%
	CMA After	96.31%	96.7%	91.84%	92.9%	86.27%	97.1%	84.52%	88.0%
3	MMA Before	96.70%	97.0%	92.93%	92.4%	87.59%	95.8%	84.89%	85.3%
	CMA Before	96.70%	97.0%	92.93%	92.4%	87.59%	95.8%	84.89%	85.3%
	MMA After	97.21%	96.8%	92.45%	92.9%	87.88%	97.0%	84.41%	84.5%
	CMA After	94.94%	97.4%	91.89%	92.9%	86.62%	96.3%	84.56%	85.3%
4	MMA Before	96.25%	96.0%	92.50%	92.9%	89.84%	96.7%	86.11%	86.2%
	CMA Before	96.25%	96.0%	92.50%	92.9%	89.84%	96.7%	86.11%	86.2%
	MMA After	96.93%	96.4%	92.26%	93.5%	87.88%	97.1%	84.80%	87.2%
	CMA After	95.71%	95.8%	90.79%	92.8%	82.91%	96.6%	83.85%	86.9%
5	MMA Before	97.29%	96.4%	92.24%	92.1%	87.42%	96.5%	86.59%	85.1%
	CMA Before	97.29%	96.4%	92.24%	92.1%	87.42%	96.5%	86.59%	85.1%
	MMA After	96.75%	97.2%	92.73%	91.9%	87.34%	97.1%	84.69%	86.2%
	CMA After	96.38%	96.5%	91.04%	91.8%	85.94%	96.6%	85.43%	85.5%
6	MMA Before	96.51%	96.8%	93.31%	92.8%	92.56%	96.7%	85.24%	84.8%
	CMA Before	96.51%	96.8%	93.31%	92.8%	92.56%	96.7%	85.24%	84.8%
	MMA After	96.51%	96.7%	92.09%	94.0%	87.05%	97.0%	85.99%	86.7%
	CMA After	96.30%	96.9%	91.46%	93.8%	87.06%	96.8%	83.70%	85.9%
7	MMA Before	96.86%	96.8%	93.13%	91.7%	87.47%	96.7%	86.66%	85.8%
	CMA Before	96.86%	96.8%	93.13%	91.7%	87.47%	96.7%	86.66%	85.8%
	MMA After	96.11%	97.4%	92.43%	93.2%	86.83%	97.1%	85.25%	85.2%
	CMA After	95.97%	96.9%	92.59%	92.0%	85.40%	96.9%	85.40%	86.6%
8	MMA Before	97.16%	96.7%	93.40%	92.8%	86.44%	96.8%	86.70%	84.6%
	CMA Before	97.16%	96.7%	93.40%	92.8%	86.44%	96.8%	86.70%	84.6%
	MMA After	96.67%	96.9%	92.96%	93.1%	87.33%	97.3%	85.63%	86.6%
	CMA After	96.43%	97.5%	92.71%	93.0%	87.16%	97.3%	85.59%	85.6%
9	MMA Before	97.51%	96.5%	92.28%	92.4%	86.60%	97.0%	86.24%	85.7%
	CMA Before	97.51%	96.5%	92.28%	92.4%	86.60%	97.0%	86.24%	85.7%
	MMA After	96.61%	96.9%	92.49%	92.7%	87.10%	97.5%	84.76%	85.9%
	CMA After	96.28%	97.1%	91.26%	92.9%	86.31%	97.4%	85.26%	86.4%

MMA: {Before/After}: Monolithic Model Accuracy {before/after} the Change, CMA: {Before/After}: Composed Model Accuracy {before/after} the Change.

Table 4: Change data shape after a model has been trained

Model	Increase						Decrease					
	Pre-MA	OMA	Po-MA	Po-TP	MNTP	MTP	Pre-MA	OMA	Po-MA	Po-TP	MNTP	MTP
MNIST-1	88.3%	91.0%	95.7%	38965	500	38465	94.6%	94.3%	87.9%	2950	38965	39200
MNIST-2	90.6%	90.0%	96.7%	41415	5400	38465	96.7%	91.4%	90.0%	5400	41415	39200
MNIST-3	93.1%	90.7%	96.6%	43865	7850	38465	97.1%	90.8%	91.4%	7850	43865	39200
MNIST-4	93.0%	90.0%	96.9%	46315	10300	38465	96.7%	91.4%	90.0%	10300	46315	39200
EMNIST-1	75.7%	76.7%	88.2%	38965	500	38465	88.5%	87.9%	75.0%	2950	38965	39200
EMNIST-2	80.9%	82.2%	91.8%	41415	5400	38465	92.1%	88.4%	82.0%	5400	41415	39200
EMNIST-3	83.5%	77.5%	92.2%	43865	7850	38465	92.2%	85.6%	84.2%	7850	43865	39200
EMNIST-4	84.2%	81.1%	92.2%	46315	10300	38465	92.1%	83.9%	86.2%	10300	46315	39200
FMNIST-1	77.1%	83.5%	85.0%	38965	500	38465	85.0%	81.8%	76.9%	2950	38965	39200
FMNIST-2	79.0%	85.7%	86.3%	41415	5400	38465	87.0%	81.6%	78.9%	5400	41415	39200
FMNIST-3	78.7%	84.1%	87.3%	43865	7850	38465	87.7%	80.8%	79.3%	7850	43865	39200
FMNIST-4	81.0%	83.8%	87.2%	46315	10300	38465	87.4%	78.1%	81.3%	10300	46315	39200
KMNIST-1	64.7%	69.1%	81.5%	38965	500	38465	80.7%	78.9%	65.7%	2950	38965	39200
KMNIST-2	72.0%	70.2%	85.0%	41415	5400	38465	86.0%	80.6%	73.1%	5400	41415	39200
KMNIST-3	75.9%	65.4%	85.6%	43865	7850	38465	85.8%	80.0%	71.5%	7850	43865	39200
KMNIST-4	78.3%	67.1%	85.8%	46315	10300	38465	84.4%	70.5%	77.7%	10300	46315	39200

Table 5: Change in Data Type

Model	CIDT <: MIDT		MIDT <: CIDT			
	Pre-MA	Po-MA	Pre-MA	Po-MA	OMA-CIDT	OMA-MIDT
MNIST-1	94.51%	94.51%	92.80%	9.90%	92.80%	92.80%
MNIST-2	96.53%	96.53%	95.28%	9.74%	95.28%	95.28%
MNIST-3	96.67%	96.67%	96.82%	9.74%	96.82%	96.82%
MNIST-4	97.02%	97.02%	97.06%	9.82%	97.06%	97.06%
EMNIST-1	80.76%	80.76%	79.99%	10.00%	80.00%	79.99%
EMNIST-2	87.30%	87.30%	89.91%	10.00%	89.88%	89.91%
EMNIST-3	91.36%	91.36%	91.54%	10.04%	91.54%	91.54%
EMNIST-4	91.98%	91.98%	92.15%	19.83%	92.15%	92.15%
FMNIST-1	81.25%	81.25%	74.51%	10.31%	74.51%	74.51%
FMNIST-2	84.59%	84.59%	85.13%	26.92%	85.13%	85.13%
FMNIST-3	85.00%	85.00%	85.05%	25.98%	85.05%	85.05%
FMNIST-4	87.25%	87.25%	86.71%	16.42%	86.71%	86.71%
KMNIST-1	70.42%	70.39%	69.18%	12.16%	69.21%	69.18%
KMNIST-2	81.27%	81.27%	80.09%	10.00%	80.15%	80.09%
KMNIST-3	81.66%	81.72%	83.33%	10.00%	83.36%	83.33%
KMNIST-4	84.50%	84.50%	85.45%	21.99%	85.43%	85.45%