



# Video Alchemist – Multi-subject Open-set Personalization in Video Generation

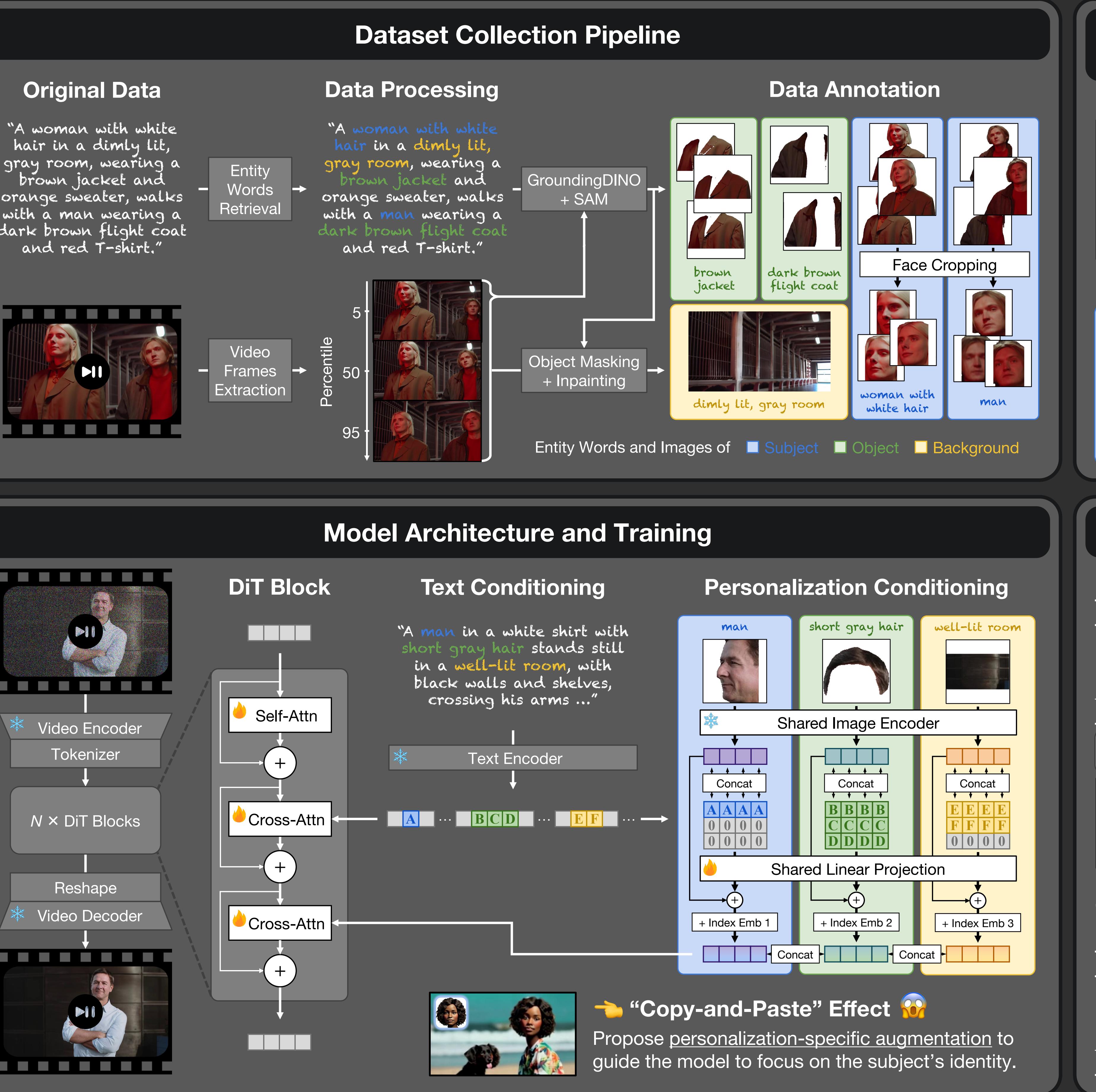
Tsai-Shien Chen<sup>1,2</sup>, Aliaksandr Siarohin<sup>1</sup>, Willi Menapace<sup>1</sup>, Yuwei Fang<sup>1</sup>, Kwot Sin Lee<sup>1</sup>, Ivan Skorokhodov<sup>1</sup>, Kfir Aberman<sup>1</sup>, Jun-Yan Zhu<sup>3</sup>, Ming-Hsuan Yang<sup>2</sup>, Sergey Tulyakov<sup>1</sup>

<sup>1</sup>Snap Inc., <sup>2</sup>UC Merced, <sup>3</sup>CMU



### Multi-subject Open-set Video Personalization

Ref Images	Generated Video

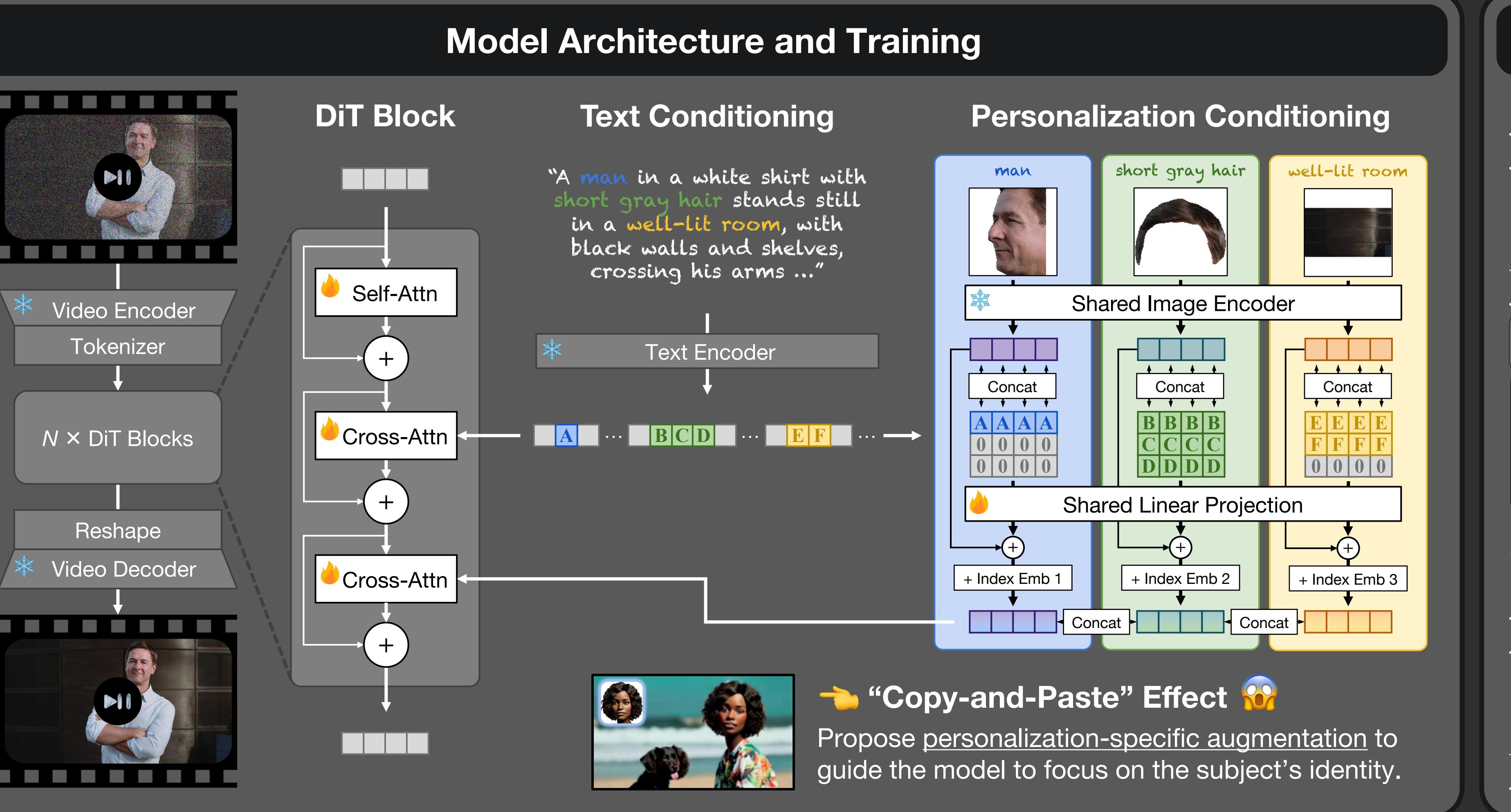


### New Benchmark 🔥 MSRVTT-Personalization

Ground Truth	Metrics
	<ul style="list-style-type: none"> <li>Text similarity</li> <li>Video similarity</li> <li>Subject similarity</li> <li>Face similarity</li> <li>Dynamic degree</li> </ul>
Personalization Annotations	

### Learn More?

Project Page
Github



👉 “Copy-and-Paste” Effect 😱  
Propose personalization-specific augmentation to guide the model to focus on the subject’s identity.

### Comparisons with the SOTAs

Open-set Subject Mode				
Method	Text-S↑	Vid-S↑	Subj-S↑	
ELITE	0.245	0.620	0.359	-
VideoBooth	0.222	0.612	0.395	0.448
DreamVideo	0.261	0.611	0.310	0.311
<b>Video Alchemist</b>	<b>0.269</b>	<b>0.732</b>	<b>0.617</b>	<b>0.466</b>

A bearded man in gray clothes brushes a brown horse [...].

Method	Text-S↑	Vid-S↑	Face-S↑	Dync↑
ELITE	0.251	0.648	0.269	-
PhotoMaker	0.278	0.569	0.189	-
Magic-Me	0.251	0.602	0.135	0.418
<b>Video Alchemist</b>	<b>0.273</b>	<b>0.687</b>	<b>0.382</b>	<b>0.424</b>

### Ablation Study

Use CLIP	No word token	No augmentation	Video Alchemist