Machine Learning for Exploring Small Polaron Configurational Space (Supplementary Information)

Viktor C. Birschitzky[1,2,3], Florian Ellinger[1,2], Ulrike Diebold[3], Michele Reticcioli[1], and Cesare Franchini[1,4]

[1] *University of Vienna, Faculty of Physics and Center for Computational Materials Science, Vienna, Austria*
[2] *University of Vienna, Vienna Doctoral School in Physics,Boltzmanngasse 5, 1090 Vienna, Austria*
[3] *Institute of Applied Physics, Technische Universität Wien, 1040 Vienna, Austria*
[4] *Department of Physics and Astronomy 'Augusto Righi', Alma Mater Studiorum - Università di Bologna, Bologna, 40127 Italy*

Supplementary Notes

Datasets. The distribution of polarons in the training databases for different V_{O} and Nb-doping concentrations are collected in Supplementary Table 1 and Supplementary Table 2, respectively.

c_{V_O}	# Configurations	$\mathrm{Ti}^{\mathrm{A}}_{\mathrm{S0}}$	$\mathrm{Ti}^{\mathrm{B}}_{\mathrm{S0}}$	$\mathrm{Ti}^{\mathrm{A}}_{\mathrm{S1}}$	$\mathrm{Ti}^{\mathrm{B}}_{\mathrm{S1}}$	$\mathrm{Ti}^{\mathrm{A}}_{\mathrm{S2}}$	$\mathrm{Ti}^{\mathrm{B}}_{\mathrm{S2}}$	$\min(\bar{E}^{\mathrm{DFT}}_{\mathrm{pol}})$	$\max(\bar{E}^{\mathrm{DFT}}_{\mathrm{pol}})$
5.5%	52	6	0	97	1	0	0	-0.3	-0.12
11.1%	86	99	0	229	13	3	0	-0.37	-0.04
16.7%	116	214	0	462	18	2	0	-0.41	-0.29
22.2%	61	192	0	239	39	18	0	-0.44	-0.36
27.8%	27	125	0	108	28	9	0	-0.43	-0.36
33.3%	11	44	16	56	16	0	0	-0.43	-0.38
38.9%	19	98	56	95	11	6	0	-0.39	-0.35
44.4%	46	219	170	261	57	9	20	-0.35	-0.28
50%	74	450	383	338	97	64	0	-0.37	-0.32
total	492	1447	625	1885	280	111	20	-0.44	-0.04

Supplementary Table 1: FPMD dataset of configurations in rutile $\mathrm{TiO_2}(110)$. Number of inequivalent polaronic configurations generated in FPMD for every defect concentration, and corresponding site-dependent polaron occupation (number of times the site is occupied by a polaron). The last columns display the ranges of mean polaronic energies in eV at each concentration (a characterization of the energy distribution generated via the ML model can be found in Supplementary Figure 9)

Number of possible configurations. We briefly discuss the number of possible configurations in the $\mathrm{TiO_2}$ super cell. In our setup, consisting of a 9×2 large unit-cell with two of the five stochiometric layers fixed to bulk positions (see Methods in the main text), the $\mathrm{TiO_2}$ slab contains 108 Ti sites that can possibly host polarons (36 sites per $S0$, $S1$ and $S2$ layer). In the simplest case of one oxygen vacancy and two excess electrons in the slab ($c_{V_{\mathrm{O}}}$= 5.5%), the number of possible polaronic configurations (with no symmetry applied) is given by the binomial coefficient, $\binom{108}{2} = 5778$. At higher concentration, the number of

c_{Nb}	# Configurations	$\mathrm{Ti}_{\mathrm{S0}}$	$\mathrm{Ti}_{\mathrm{S1}}$	$\mathrm{Ti}_{\mathrm{S2}}$	$\min(\bar{E}_{\mathrm{pol}}^{\mathrm{DFT}})$	$\max(\bar{E}_{\mathrm{pol}}^{\mathrm{DFT}})$
3.3%	95	146	174	60	-0.39	-0.26
4.2%	79	153	73	169	-0.38	-0.2
5.0%	42	102	50	100	-0.4	-0.26
5.8%	74	178	185	155	-0.43	-0.27
6.7%	89	254	236	222	-0.43	-0.33
total	379	833	718	706	-0.43	-0.2

Supplementary Table 2: Randomly generated dataset of configurations in $\mathrm{SrTiO}_3(001)$. Number of inequivalent polaronic configurations generated for every defect concentration, and corresponding site-dependent polaron occupation (number of times the site is occupied by a polaron). The last columns display the ranges of mean polaronic energies in eV at each concentration.

configurations explodes, *e.g.*, for 4 oxygen vacancies we obtain $\binom{108}{8} \approx 3 \cdot 10^{11}$ polaronic configurations. We note that, especially at high defect concentration, exploiting symmetry operations to simplify the problem does not bring any considerable advantage.

Supplementary Methods

Occupation matrices. We used the Occupation Matrix Control scheme as developed by Allen et al. (2014) to localize polarons at specific sites in a two step protocol. In the first step, occupation matrices of chosen Ti sites are fixed in order to ensure polaron localization at the specified sites. The second step consists of an unrestricted relaxation, which starts from the previously determined distorted structure. The second step is crucial to obtain reliable self-consistent results of the polaron trapping.

In the following, we list the spin-resolved density matrices used in the first step of the relaxation procedure. In the case of rutile $\mathrm{TiO_2}(110)$ we employed specific occupation matrices for every site type ($\mathrm{Ti}^{\mathrm{A}}_{\mathrm{S0}}$, $\mathrm{Ti}^{\mathrm{A}}_{\mathrm{S1}}$, $\mathrm{Ti}^{\mathrm{B}}_{\mathrm{S1}}$, $\mathrm{Ti}^{\mathrm{B}}_{\mathrm{S0}}$), here listed with the corresponding spin-channel ↑ and ↓ arrows. All occupation matrices have been extracted from a single polaron configuration at $c_{V_{\mathrm{O}}}$=38.9%. Polarons at $\mathrm{Ti}^{\mathrm{A}}_{\mathrm{S0}}$, $\mathrm{Ti}^{\mathrm{A}}_{\mathrm{S1}}$, and $\mathrm{Ti}^{\mathrm{B}}_{\mathrm{S1}}$ showed well defined orbital characters, while the unstable $\mathrm{Ti}^{\mathrm{B}}_{\mathrm{S0}}$ polarons varied in their specific orbital character. Below, the used input occupation matrices are listed:

$$\mathrm{Ti}^{\mathrm{A}}_{\mathrm{S0}},\uparrow=\begin{bmatrix} 0.32 & -0.01 & -0.03 & -0.01 & 0.01 \\ -0.01 & 0.57 & -0.06 & -0.38 & -0.19 \\ -0.03 & -0.06 & 0.19 & 0.04 & -0.01 \\ -0.01 & -0.38 & 0.04 & 0.40 & 0.15 \\ 0.01 & -0.19 & -0.01 & 0.15 & 0.15 \end{bmatrix} \quad \mathrm{Ti}^{\mathrm{A}}_{\mathrm{S0}},\downarrow=\begin{bmatrix} 0.30 & -0.00 & -0.03 & -0.00 & 0.01 \\ -0.00 & 0.06 & -0.00 & 0.02 & 0.00 \\ -0.03 & -0.00 & 0.18 & -0.00 & -0.03 \\ -0.00 & 0.02 & -0.00 & 0.08 & -0.00 \\ 0.01 & 0.00 & -0.03 & -0.00 & 0.08 \end{bmatrix}$$

$$\mathrm{Ti}^{\mathrm{A}}_{\mathrm{S1}},\uparrow=\begin{bmatrix} 0.09 & 0.00 & -0.02 & -0.00 & 0.01 \\ 0.00 & 0.09 & -0.00 & -0.00 & -0.00 \\ -0.02 & -0.00 & 0.76 & 0.00 & -0.23 \\ -0.00 & -0.00 & 0.00 & 0.28 & -0.00 \\ 0.01 & -0.00 & -0.23 & -0.00 & 0.43 \end{bmatrix} \quad \mathrm{Ti}^{\mathrm{A}}_{\mathrm{S1}},\downarrow=\begin{bmatrix} 0.08 & 0.00 & -0.00 & -0.00 & -0.00 \\ 0.00 & 0.08 & -0.00 & -0.00 & -0.00 \\ -0.00 & -0.00 & 0.10 & -0.00 & 0.09 \\ -0.00 & -0.00 & -0.00 & 0.26 & 0.00 \\ -0.00 & -0.00 & 0.09 & 0.00 & 0.24 \end{bmatrix}$$

$$\mathrm{Ti}^{\mathrm{B}}_{\mathrm{S1}},\uparrow=\begin{bmatrix} 0.29 & -0.00 & -0.00 & 0.00 & -0.01 \\ -0.00 & 0.08 & -0.00 & 0.00 & -0.01 \\ -0.00 & -0.00 & 0.29 & 0.00 & 0.06 \\ 0.00 & 0.00 & 0.00 & 0.09 & -0.02 \\ -0.01 & -0.01 & 0.06 & -0.02 & 0.88 \end{bmatrix} \quad \mathrm{Ti}^{\mathrm{B}}_{\mathrm{S1}},\downarrow=\begin{bmatrix} 0.28 & -0.00 & -0.00 & 0.00 & 0.00 \\ -0.00 & 0.07 & -0.00 & 0.00 & 0.00 \\ -0.00 & -0.00 & 0.26 & 0.00 & -0.02 \\ 0.00 & 0.00 & 0.00 & 0.08 & -0.00 \\ 0.00 & 0.00 & -0.02 & -0.00 & 0.06 \end{bmatrix}$$

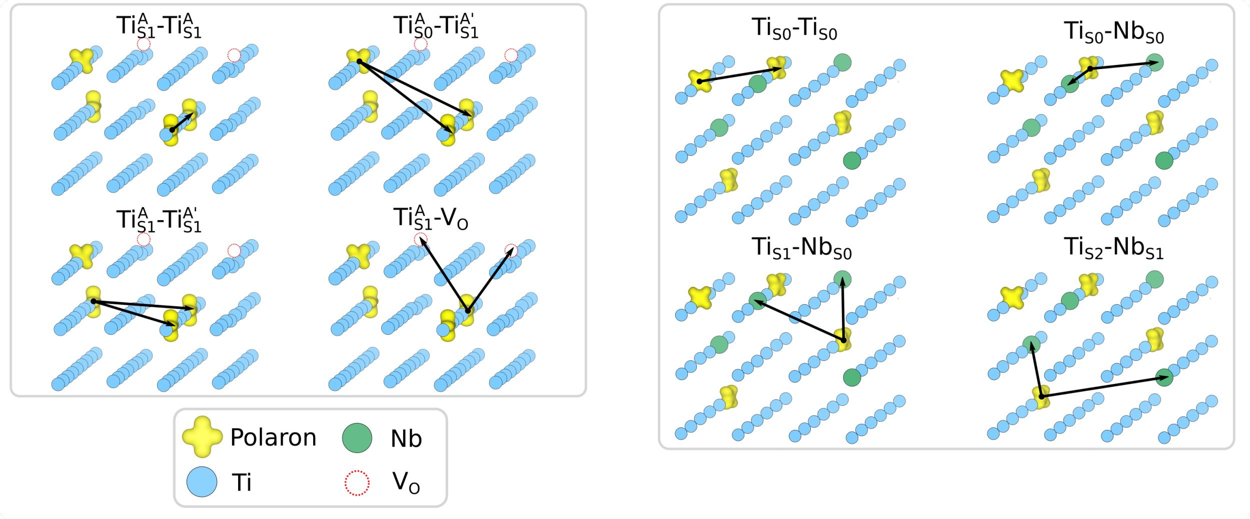
$$\mathrm{Ti}^{\mathrm{B}}_{\mathrm{S0}},\uparrow=\begin{bmatrix} 0.11 & -0.00 & -0.05 & 0.03 & 0.05 \\ -0.00 & 0.10 & -0.01 & 0.00 & 0.03 \\ -0.05 & -0.01 & 0.58 & -0.21 & -0.24 \\ 0.03 & 0.00 & -0.21 & 0.35 & 0.09 \\ 0.05 & 0.03 & -0.24 & 0.09 & 0.49 \end{bmatrix} \quad \mathrm{Ti}^{\mathrm{B}}_{\mathrm{S0}},\downarrow=\begin{bmatrix} 0.09 & -0.01 & 0.01 & 0.00 & 0.01 \\ -0.01 & 0.08 & 0.00 & -0.00 & 0.01 \\ 0.01 & 0.00 & 0.11 & 0.01 & 0.08 \\ 0.00 & -0.00 & 0.01 & 0.24 & -0.06 \\ 0.01 & 0.01 & 0.08 & -0.06 & 0.25 \end{bmatrix}$$

$$\mathrm{Ti}^{\mathrm{B}}_{\mathrm{S0}},\uparrow=\begin{bmatrix} 0.54 & 0.41 & 0.03 & 0.00 & 0.02 \\ 0.41 & 0.48 & 0.03 & 0.00 & 0.02 \\ 0.03 & 0.03 & 0.13 & -0.03 & 0.03 \\ 0.00 & 0.00 & -0.03 & 0.32 & -0.06 \\ 0.02 & 0.02 & 0.03 & -0.06 & 0.24 \end{bmatrix} \quad \mathrm{Ti}^{\mathrm{B}}_{\mathrm{S0}},\downarrow=\begin{bmatrix} 0.07 & -0.02 & -0.00 & -0.00 & -0.01 \\ -0.02 & 0.06 & -0.00 & 0.00 & -0.01 \\ -0.00 & -0.00 & 0.10 & -0.01 & 0.05 \\ -0.00 & 0.00 & -0.01 & 0.26 & -0.08 \\ -0.01 & -0.01 & 0.05 & -0.08 & 0.21 \end{bmatrix}$$

For $\mathrm{SrTiO_3}(001)$, we employed the same occupation matrix for all localization sites in both the generation of the reference database and in the exhaustive search:

$$\mathrm{Ti},\uparrow=\begin{bmatrix} 0.28 & -0.00 & -0.02 & 0.00 & -0.07 \\ -0.00 & 0.41 & -0.00 & -0.40 & 0.00 \\ -0.02 & -0.00 & 0.24 & 0.00 & 0.00 \\ 0.00 & -0.40 & 0.00 & 0.59 & -0.00 \\ -0.07 & 0.00 & 0.00 & -0.00 & 0.12 \end{bmatrix} \quad \mathrm{Ti},\downarrow=\begin{bmatrix} 0.27 & 0.00 & -0.02 & 0.00 & -0.07 \\ 0.00 & 0.07 & -0.00 & 0.01 & -0.00 \\ -0.02 & -0.00 & 0.23 & -0.00 & 0.00 \\ 0.00 & 0.01 & -0.00 & 0.06 & 0.00 \\ -0.07 & -0.00 & 0.00 & 0.00 & 0.11 \end{bmatrix}$$

Descriptors. In this section we briefly describe the most important concepts employed for constructing our descriptor.



Supplementary Figure 1: (left) Four of the possible 60 interaction categories in a rutile $TiO_2(110)$ polaron configuration at c_{V_O}= 11.1% are shown. For clarity, O-atoms are hidden and the locations of oxygen vacancies are marked. Possible polaron hosting sites are shown in blue and polarons are represented by charge densities taken from a DFT-calculation. The starting point of arrows indicates the polaron, whose descriptor ought to be calculated and the arrow indicates interactions with polarons/defects at specific sites (vectors connecting the considered sites are not necessarily the shortest distance in the periodic cell, although in practice the minimum image convention was used to construct descriptors). (right) Simlilarly to the left hand side, some examples of interaction categories in $SrTiO_3(001)$ are depicted. Again, O- and Sr-atoms have been omitted for clarity, but Nb-dopants are displayed.

Interaction Categories. To ensure consistent representations of the polaron environment we structure the 2-body-interaction terms used to construct the descriptor via interaction categories. An interaction category depends on the type of polaron-hosting site, and on the nature of interaction (whether with another polaron of with a donor defect). Supplementary Figure 1 collects examples of interaction categories for the two materials of our study. The arrows in the left side of Supplementary Figure 1 show four interaction categories for rutile $TiO_2(110)$: polaron-polaron on the same-row ($\mathrm{Ti}^{\mathrm{A}}_{\mathrm{S1}}$-$\mathrm{Ti}^{\mathrm{A}}_{\mathrm{S1}}$), non-stacked on different layers ($\mathrm{Ti}^{\mathrm{A}}_{\mathrm{S0}}$-$\mathrm{Ti}^{\mathrm{A}'}_{\mathrm{S1}}$), non-stacked on same layer ($\mathrm{Ti}^{\mathrm{A}}_{\mathrm{S1}}$-$\mathrm{Ti}^{\mathrm{A}'}_{\mathrm{S1}}$), and polaron-defect ($\mathrm{Ti}^{\mathrm{A}}_{\mathrm{S1}}$-$V_{\mathrm{O}}$). The right panel similarly shows four possible interaction categories in $SrTiO_3$: polaron-polaron on the same layer ($\mathrm{Ti}_{\mathrm{S0}}$-$\mathrm{Ti}_{\mathrm{S0}}$), polaron-defect on the same ($\mathrm{Ti}_{\mathrm{S0}}$-$\mathrm{Nb}_{\mathrm{S0}}$) and different ($\mathrm{Ti}_{\mathrm{S1}}$-$\mathrm{Nb}_{\mathrm{S0}}$, $\mathrm{Ti}_{\mathrm{S2}}$-$\mathrm{Nb}_{\mathrm{S1}}$) layers. A full list of possible interaction categories for $TiO_2(110)$ and $SrTiO_3(001)$ is provided in Supplementary Table 3 and Supplementary Table 4, respectively.

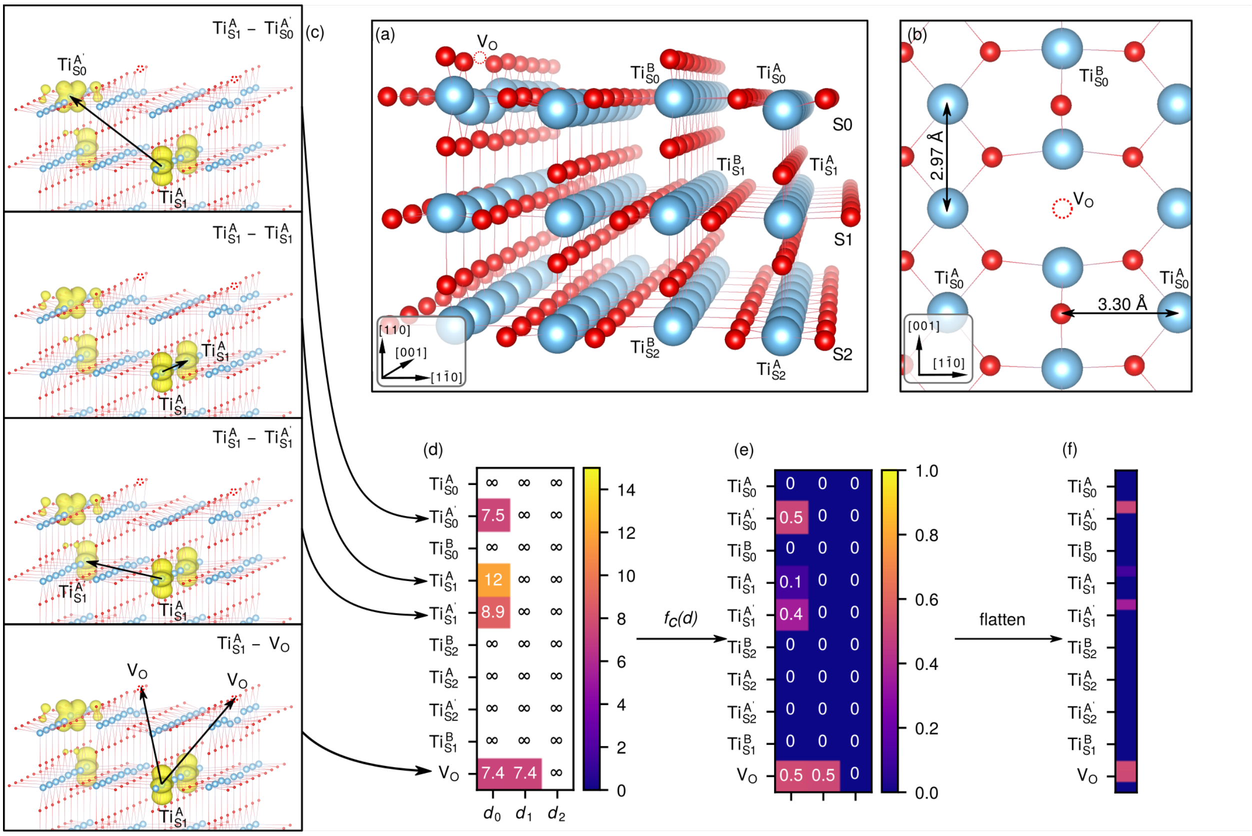
Layers	Interaction type	Trapping site					
	Ti−Ti or Ti−V_{O}	$\mathrm{Ti^{A}_{S0}}$	$\mathrm{Ti^{B}_{S0}}$	$\mathrm{Ti^{A}_{S1}}$	$\mathrm{Ti^{B}_{S1}}$	$\mathrm{Ti^{A}_{S2}}$	$\mathrm{Ti^{B}_{S2}}$
Sl − S0	A − A/B − B (stacked)	$\mathrm{Ti^{A}_{S0}}$-$\mathrm{Ti^{A}_{S0}}$	$\mathrm{Ti^{B}_{S0}}$-$\mathrm{Ti^{B}_{S0}}$	$\mathrm{Ti^{A}_{S1}}$-$\mathrm{Ti^{A}_{S0}}$	$\mathrm{Ti^{B}_{S1}}$-$\mathrm{Ti^{B}_{S0}}$	$\mathrm{Ti^{A}_{S2}}$-$\mathrm{Ti^{A}_{S0}}$	$\mathrm{Ti^{B}_{S2}}$-$\mathrm{Ti^{B}_{S0}}$
	A − A′/B − B′ (non-stacked)	$\mathrm{Ti^{A}_{S0}}$-$\mathrm{Ti^{A'}_{S0}}$	$\mathrm{Ti^{B}_{S0}}$-$\mathrm{Ti^{B'}_{S0}}$	$\mathrm{Ti^{A}_{S1}}$-$\mathrm{Ti^{A'}_{S0}}$	$\mathrm{Ti^{B}_{S1}}$-$\mathrm{Ti^{B'}_{S0}}$	$\mathrm{Ti^{A}_{S2}}$-$\mathrm{Ti^{A'}_{S0}}$	$\mathrm{Ti^{B}_{S2}}$-$\mathrm{Ti^{B'}_{S0}}$
	A-B (different coordination)	$\mathrm{Ti^{A}_{S0}}$-$\mathrm{Ti^{B}_{S0}}$	$\mathrm{Ti^{B}_{S0}}$-$\mathrm{Ti^{A}_{S0}}$	$\mathrm{Ti^{A}_{S1}}$-$\mathrm{Ti^{B}_{S0}}$	$\mathrm{Ti^{B}_{S1}}$-$\mathrm{Ti^{A}_{S0}}$	$\mathrm{Ti^{A}_{S2}}$-$\mathrm{Ti^{B}_{S0}}$	$\mathrm{Ti^{B}_{S2}}$-$\mathrm{Ti^{A}_{S0}}$
Sl − S1	A-A/B-B	$\mathrm{Ti^{A}_{S0}}$-$\mathrm{Ti^{A}_{S1}}$	$\mathrm{Ti^{B}_{S0}}$-$\mathrm{Ti^{B}_{S1}}$	$\mathrm{Ti^{A}_{S1}}$-$\mathrm{Ti^{A}_{S1}}$	$\mathrm{Ti^{B}_{S1}}$-$\mathrm{Ti^{B}_{S1}}$	$\mathrm{Ti^{A}_{S2}}$-$\mathrm{Ti^{A}_{S1}}$	$\mathrm{Ti^{B}_{S2}}$-$\mathrm{Ti^{B}_{S1}}$
	A-A'/B-B'	$\mathrm{Ti^{A}_{S0}}$-$\mathrm{Ti^{A'}_{S1}}$	$\mathrm{Ti^{B}_{S0}}$-$\mathrm{Ti^{B'}_{S1}}$	$\mathrm{Ti^{A}_{S1}}$-$\mathrm{Ti^{A'}_{S1}}$	$\mathrm{Ti^{B}_{S1}}$-$\mathrm{Ti^{B'}_{S1}}$	$\mathrm{Ti^{A}_{S2}}$-$\mathrm{Ti^{A'}_{S1}}$	$\mathrm{Ti^{B}_{S2}}$-$\mathrm{Ti^{B'}_{S1}}$
	A-B	$\mathrm{Ti^{A}_{S0}}$-$\mathrm{Ti^{B}_{S1}}$	$\mathrm{Ti^{B}_{S0}}$-$\mathrm{Ti^{A}_{S1}}$	$\mathrm{Ti^{A}_{S1}}$-$\mathrm{Ti^{B}_{S1}}$	$\mathrm{Ti^{B}_{S1}}$-$\mathrm{Ti^{A}_{S1}}$	$\mathrm{Ti^{A}_{S2}}$-$\mathrm{Ti^{B}_{S1}}$	$\mathrm{Ti^{B}_{S2}}$-$\mathrm{Ti^{A}_{S1}}$
Sl − S2	A-A/B-B	$\mathrm{Ti^{A}_{S0}}$-$\mathrm{Ti^{A}_{S2}}$	$\mathrm{Ti^{B}_{S0}}$-$\mathrm{Ti^{B}_{S2}}$	$\mathrm{Ti^{A}_{S1}}$-$\mathrm{Ti^{A}_{S2}}$	$\mathrm{Ti^{B}_{S1}}$-$\mathrm{Ti^{B}_{S2}}$	$\mathrm{Ti^{A}_{S2}}$-$\mathrm{Ti^{A}_{S2}}$	$\mathrm{Ti^{B}_{S2}}$-$\mathrm{Ti^{B}_{S2}}$
	A-A'/B-B'	$\mathrm{Ti^{A}_{S0}}$-$\mathrm{Ti^{A'}_{S2}}$	$\mathrm{Ti^{B}_{S0}}$-$\mathrm{Ti^{B'}_{S2}}$	$\mathrm{Ti^{A}_{S1}}$-$\mathrm{Ti^{A'}_{S2}}$	$\mathrm{Ti^{B}_{S1}}$-$\mathrm{Ti^{B'}_{S2}}$	$\mathrm{Ti^{A}_{S2}}$-$\mathrm{Ti^{A'}_{S2}}$	$\mathrm{Ti^{B}_{S2}}$-$\mathrm{Ti^{B'}_{S2}}$
	A-B	$\mathrm{Ti^{A}_{S0}}$-$\mathrm{Ti^{B}_{S2}}$	$\mathrm{Ti^{B}_{S0}}$-$\mathrm{Ti^{A}_{S2}}$	$\mathrm{Ti^{A}_{S1}}$-$\mathrm{Ti^{B}_{S2}}$	$\mathrm{Ti^{B}_{S1}}$-$\mathrm{Ti^{A}_{S2}}$	$\mathrm{Ti^{A}_{S2}}$-$\mathrm{Ti^{B}_{S2}}$	$\mathrm{Ti^{B}_{S2}}$-$\mathrm{Ti^{A}_{S2}}$
Sl − S0	A/B−V_{O}	$\mathrm{Ti^{A}_{S0}}$-V_{O}	$\mathrm{Ti^{B}_{S0}}$-V_{O}	$\mathrm{Ti^{A}_{S1}}$-V_{O}	$\mathrm{Ti^{B}_{S1}}$-V_{O}	$\mathrm{Ti^{A}_{S2}}$-V_{O}	$\mathrm{Ti^{B}_{S2}}$-V_{O}

Supplementary Table 3: Labeling scheme for the descriptors adopted in the ML model for rutile $\mathrm{TiO_2}$(110). Each column reports the interaction categories for a polaron localized at a specific trapping site. Rows show the trapping sites of the interacting polaron or defect (V_{O}), including a differentiation between stacked, non-stacked, and differently coordinated (*e.g.*, A-A, A-A′, A-B, respectively) interacting sites.

Layers	Interaction type	Trapping site		
	Ti−Ti or Ti−Nb	$\mathrm{Ti_{S0}}$	$\mathrm{Ti_{S1}}$	$\mathrm{Ti_{S2}}$
Sl − S0	Ti-Ti	$\mathrm{Ti_{S0}}$-$\mathrm{Ti_{S0}}$	$\mathrm{Ti_{S1}}$-$\mathrm{Ti_{S0}}$	$\mathrm{Ti_{S2}}$-$\mathrm{Ti_{S0}}$
	Ti-Nb	$\mathrm{Ti_{S0}}$-$\mathrm{Nb_{S0}}$	$\mathrm{Ti_{S1}}$-$\mathrm{Nb_{S0}}$	$\mathrm{Ti_{S2}}$-$\mathrm{Nb_{S0}}$
Sl − S1	Ti-Ti	$\mathrm{Ti_{S0}}$-$\mathrm{Ti_{S1}}$	$\mathrm{Ti_{S1}}$-$\mathrm{Ti_{S1}}$	$\mathrm{Ti_{S2}}$-$\mathrm{Ti_{S1}}$
	Ti-Nb	$\mathrm{Ti_{S0}}$-$\mathrm{Nb_{S1}}$	$\mathrm{Ti_{S1}}$-$\mathrm{Nb_{S1}}$	$\mathrm{Ti_{S2}}$-$\mathrm{Nb_{S1}}$
Sl − S2	Ti-Ti	$\mathrm{Ti_{S0}}$-$\mathrm{Ti_{S2}}$	$\mathrm{Ti_{S1}}$-$\mathrm{Ti_{S2}}$	$\mathrm{Ti_{S2}}$-$\mathrm{Ti_{S2}}$

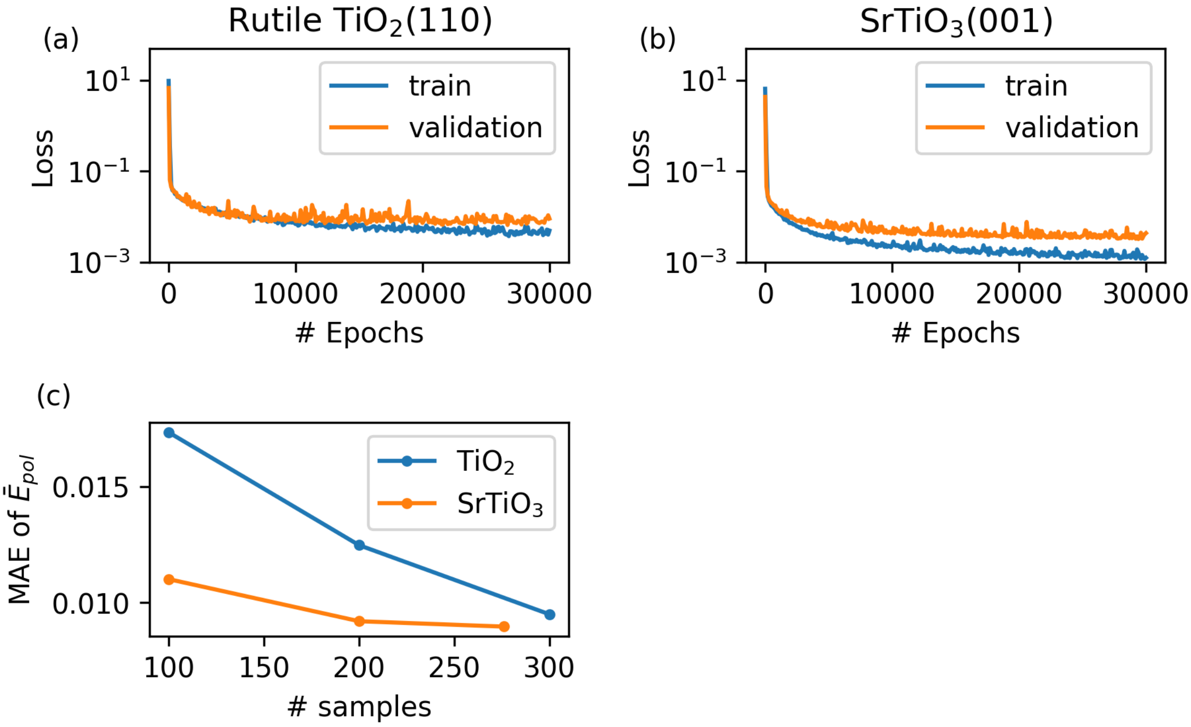
Supplementary Table 4: Labeling scheme for the descriptors adopted in the ML model for $\mathrm{SrTiO_3}$(001). Each column reports the interaction categories for a polaron localized at a specific trapping site. Rows show the trapping sites of the interacting polaron or defect (Nb).

Workflow. Supplementary Figure 2 shows the workflow for determining a polaron descriptor, considering a specific polaron configuration on $\mathrm{TiO_2}$(110). First, we recall the structural unit of rutile $\mathrm{TiO_2}$(110) and the corresponding labeling in Supplementary Figure 2(a-b). Supplementary Figure 2(c) shows a polaron configuration and all active interaction categories for the central polaron located at a $\mathrm{Ti^{A}_{S1}}$-site. In Supplementary Figure 2(d), the corresponding 2-body-interactions are determined and structured according to the interaction categories. After rescaling (Supplementary Figure 2(e)) the array is flattened (Supplementary Figure 2(f)) into a vector for further computations.



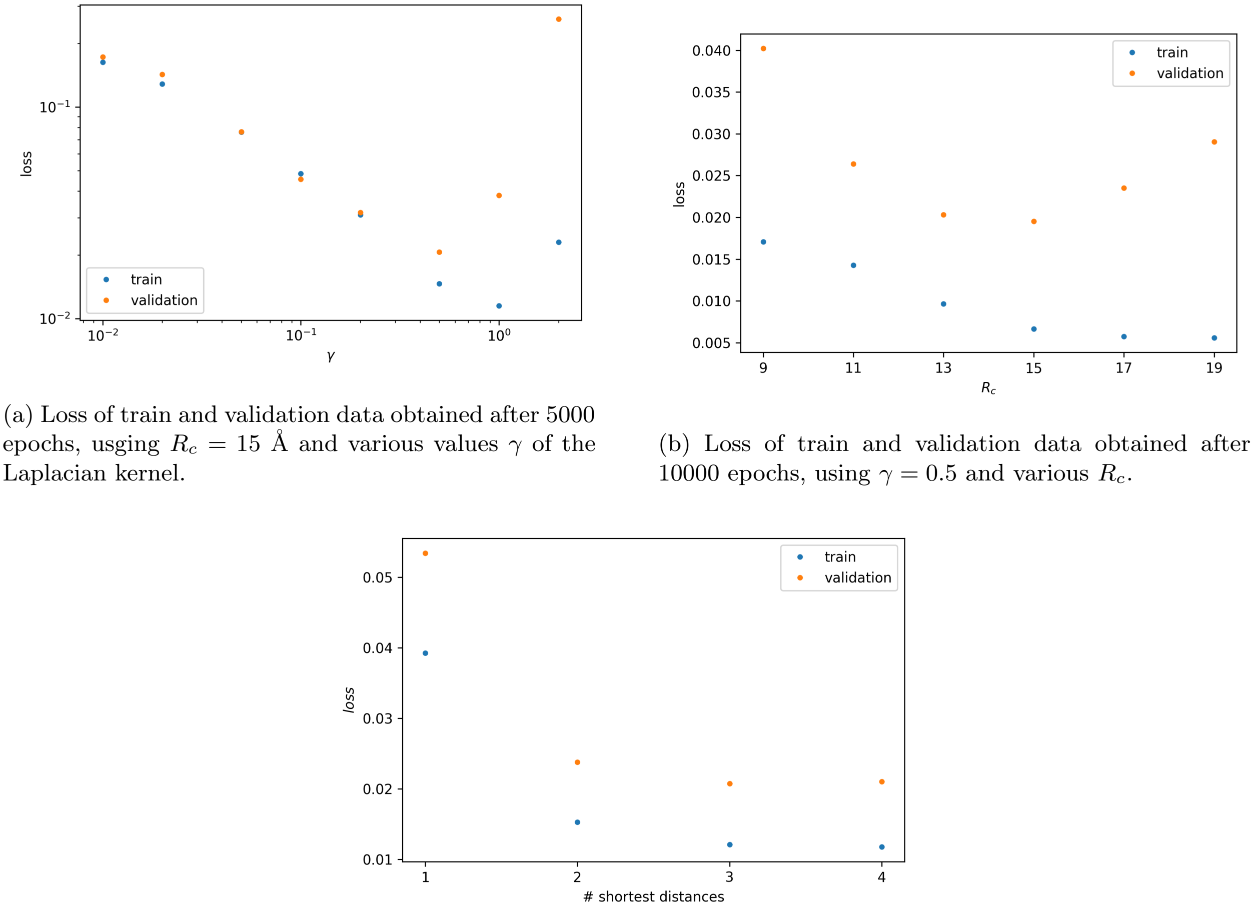
Supplementary Figure 2: (a-b) Side and top view of the rutile $TiO_2(110)$ structure with labeling of differently coordinated sites, distances and layers. Titanium sites are shown in blue and oxygen atoms in red. A dotted red circle shows the location of a V_O. (c-f) The process of calculating the descriptor vector of a Ti^A_{S1}-polaron in a configuration with four polarons and two V_O (c_{V_O}=11.1%). (c) Isosurfaces of the polaron charge (yellow). Each subpanel shows a different interaction category (see labels upper right) where the target distances are indicated by an arrow. (d) 10×3 array indicating the interacting polaron-polaron and polaron-V_O pairs for a Ti^A_{S1} polaron according to the interactions shown in panel (c). For each interaction the polaron-polaron and polaron-V_O distance d between the target and paired polaron/V_O is indicated (in Å, also displayed with a color gradient as defined in the lateral color bar.) For each interaction category only the three shortest distances (labeled d_1, d_2, and d_3) within a cutoff radius $R_c = 15$ Å are considered. If too few distances are available (generally the case for low concentrations, *i.e.*, few polarons), the distances are padded with a value greater than R_c (here ∞). (e) Illustration of the action of the rescaling function $f_c(d)$ (Equation 1 in the main text) applied to the array of distances. (f) Finally, the array is flattened into the descriptor vector $\mathbf{D}$ of the considered polaron, used as input in our machine learning model.

Training. We recorded the adapted loss (see Equation 3 in the main text) at an interval of 100 stochastic optimization epochs for the training and validation phases, using a randomized split of the databases and observed the convergence of the mean absolute error (MAE)in dependence of the number of training samples for both materials.



Supplementary Figure 3: The adapted loss function is shown for rutile TiO$_2$ (a) and SrTiO$_3$ (b) in dependence of the number of trained epochs in the stochastic parameter optimization. Panel (c) shows the convergence of the MAE of the mean polaronic energy as a function of the number of training samples.

Hyperparameter optimization. We determined optimal hyperparameters of the model (kernel parameter γ, cutoff-radius R_c and number of elements per interaction category) by optimizing the validation accuracy. The results of this procedure are displayed for rutile $TiO_2(110)$ in Supplementary Figure 4.



(a) Loss of train and validation data obtained after 5000 epochs, usging $R_c = 15$ Å and various values γ of the Laplacian kernel.

(b) Loss of train and validation data obtained after 10000 epochs, using $\gamma = 0.5$ and various R_c.

(c) Loss of train and validation data obtained after 5000 epochs for $\gamma = 0.5$, by varying the number of shortest distances included in each interaction category. While different interaction categories could be tuned separately to further optimize the descriptor, here we limited the optimization to an equal number of features per interaction category.

Supplementary Figure 4: Examples of the hyperparameter optimization based of results from oxygen defective rutile $TiO_2(110)$.

Supplementary Discussion

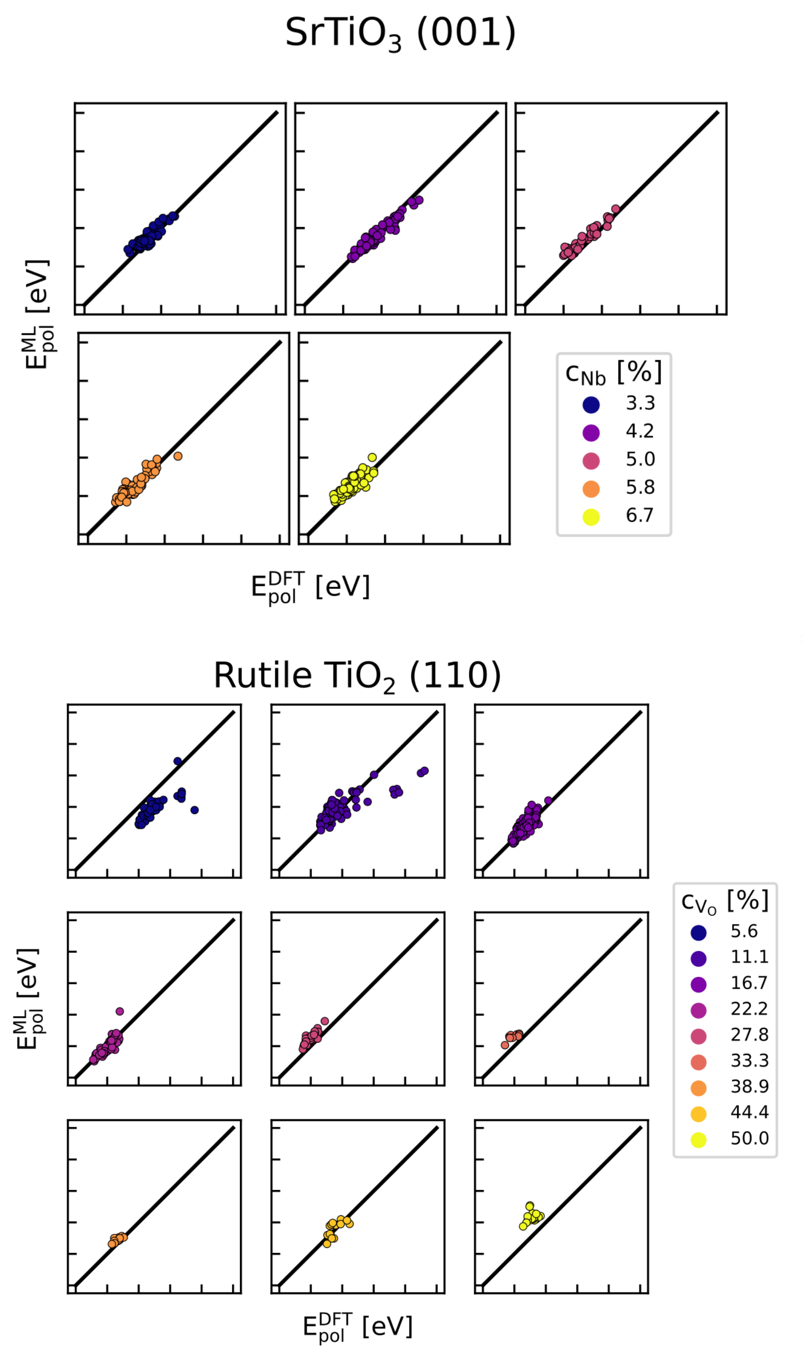
Omitted defect concentrations. Considering that the goal of this work is the efficient exploration of the polaron configurational space, the ML-model must be robust and guarantee accurate predictions for samples not visited in the training phase. Usually, the application of ML-models for extrapolation problems results in unreliable predictions, and careful tests are required. To this aim, we adopted a validation scheme, the "omitted defect concentrations" presented in the main text. Supplementary Figure 5 shows the results at every defect concentration for both the $SrTiO_3$ and TiO_2 systems. Furthermore, in Supplementary Table 5 and Supplementary Table 6, we report the mean squared error of the target quantity $\bar{E}_{\mathrm{pol}}$ as obtained in the training (including all concentrations in the database except the omitted one) and the test phase (omitted concentration only). The results show the good level of accuracy of the proposed ML model in addressing previously unknown polaron configurations. The error obtained for the omitted concentration is indeed slightly larger than the value obtained for the standard test on randomly split training-vs-test databases (see last line in the Tables).

c_{V_O}	Train	Test
5.5%	$1.11 \cdot 10^{-4}$	$2.61 \cdot 10^{-3}$
11.1%	$1.21 \cdot 10^{-4}$	$1.72 \cdot 10^{-3}$
16.7%	$1.64 \cdot 10^{-4}$	$4.85 \cdot 10^{-4}$
22.2%	$2.06 \cdot 10^{-4}$	$2.65 \cdot 10^{-4}$
27.8%	$2.43 \cdot 10^{-4}$	$5.86 \cdot 10^{-4}$
33.3%	$1.71 \cdot 10^{-4}$	$1.45 \cdot 10^{-3}$
38.9%	$1.97 \cdot 10^{-4}$	$5.32 \cdot 10^{-4}$
44.4%	$2.07 \cdot 10^{-4}$	$5.27 \cdot 10^{-4}$
50%	$1.74 \cdot 10^{-4}$	$3.04 \cdot 10^{-3}$
Randomized	$1.16 \cdot 10^{-4}$	$1.31 \cdot 10^{-4}$

Supplementary Table 5: The mean squared error of the mean polaronic energy for different test cases in the MD-dataset, reported for the training and test sets. The training was performed on configurations from eight defect concentrations, and the testing on the remaining one (labeled in column c_{V_O}). The last line shows results from a randomized split of the data from all defect concentrations.

c_{Nb}	Train	Test
3.3%	$3.01 \cdot 10^{-5}$	$2.19 \cdot 10^{-4}$
4.2%	$4.26 \cdot 10^{-5}$	$1.64 \cdot 10^{-4}$
5.0%	$3.25 \cdot 10^{-5}$	$3.79 \cdot 10^{-4}$
5.8%	$3.64 \cdot 10^{-5}$	$1.75 \cdot 10^{-4}$
6.7%	$2.92 \cdot 10^{-5}$	$4.07 \cdot 10^{-4}$
Randomized	$5.75 \cdot 10^{-5}$	$9.35 \cdot 10^{-5}$

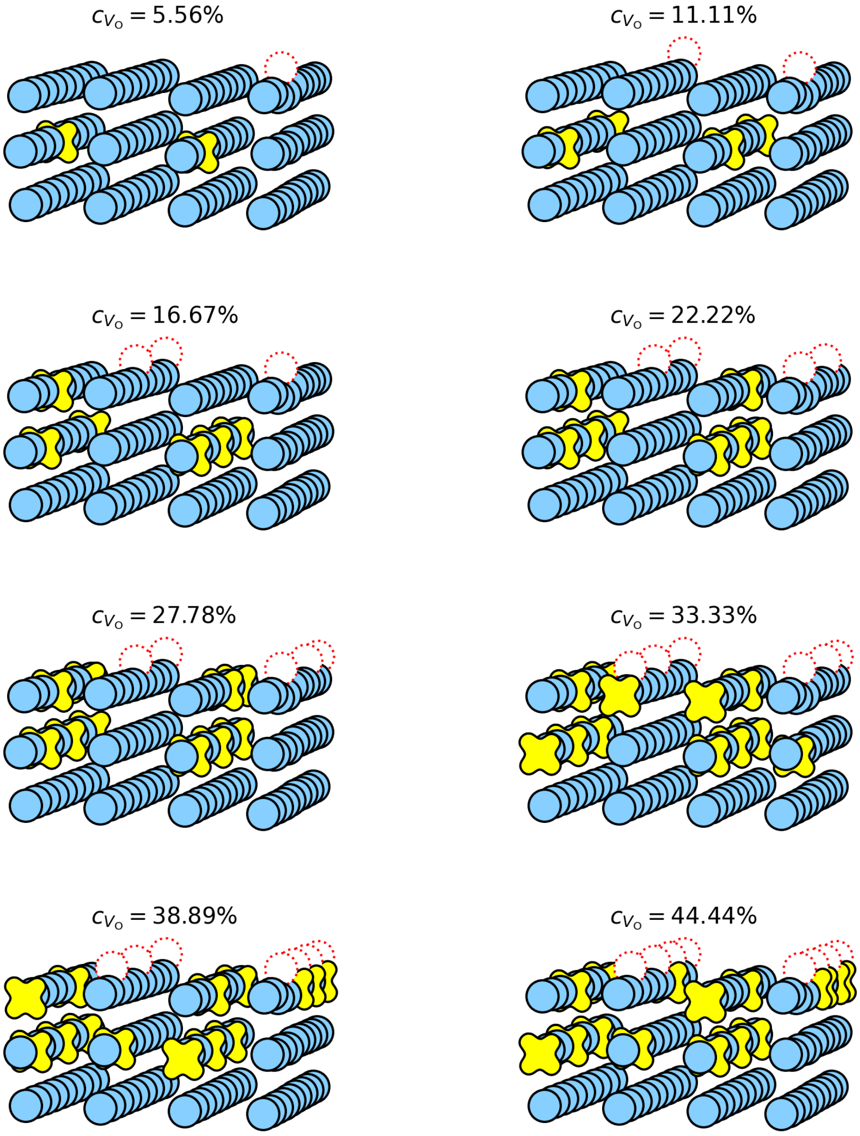
Supplementary Table 6: The mean squared error of the mean polaronic energy for different test cases in the randomized dataset, reported for the training and test sets. The training was performed on configurations from 4 defect concentrations, and the testing on the remaining one (labeled in column c_{Nb}). The last line shows results from a randomized split of the data from all defect concentrations.



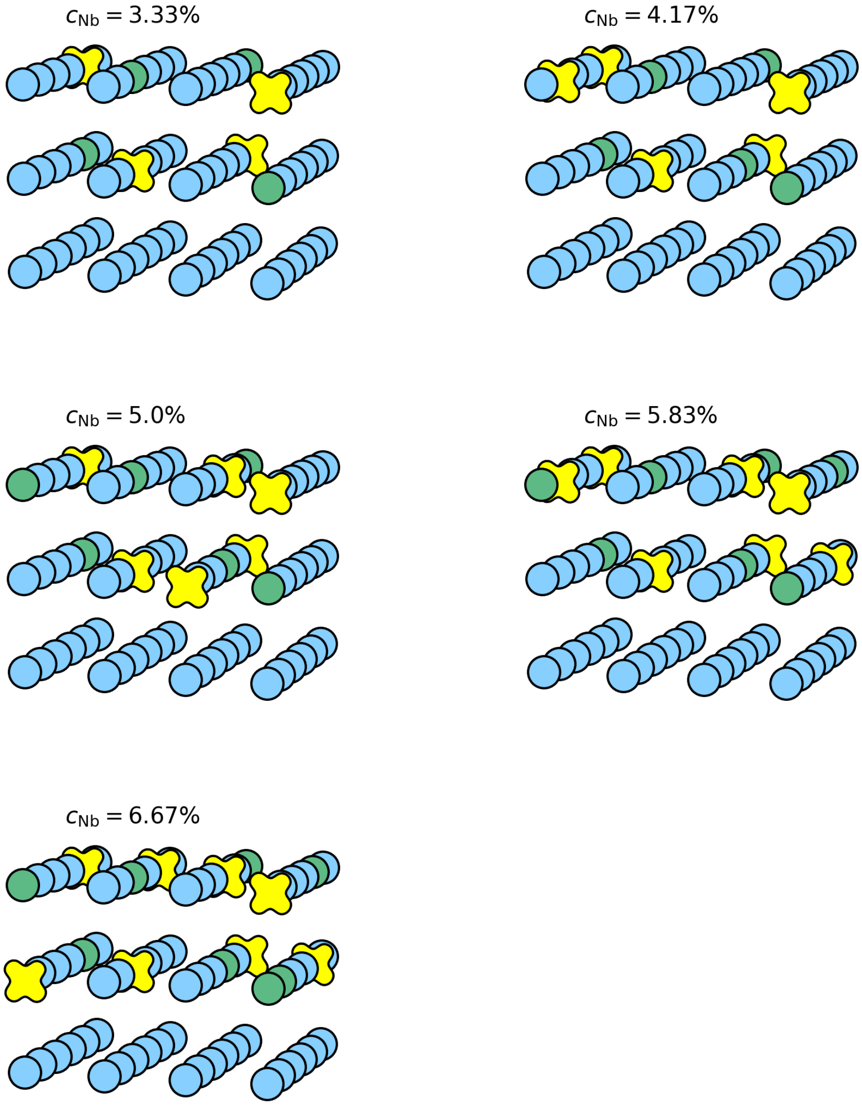
Supplementary Figure 5: The results of the validation of interpolation to omitted defect concentrations are shown for $SrTiO_3$ and TiO_2. Each panel displays only the results of the testing data from the omitted defect concentration, while the model was trained on data of all remaining defect concentrations. The color of the points encodes the defect concentration of the corresponding data.

Exhaustive Search. Supplementary Figure 6 and Supplementary Figure 7 collect the most-stable polaronic configuration as found by the machine-learning exhaustive search for every defect concentration in TiO_2 and $SrTiO_3$, respectively. These ground-state configurations were individuated by the ML algorithm among a pool of ~ 4 and ~ 2 million polaron patterns, respectively. We note that searches based purely on DFT calculations are limited to smaller datasets due to the higher computational costs: The characterization of just ~ 500 distinct polaronic configurations in TiO_2 (*i.e.*, calculations including MD simulations plus ionic relaxations for the calculation of $E_{\mathrm{pol}}^{\mathrm{DFT}}$) required us $\sim 2 \cdot 10^6$ core hours on our high-performance computing facility (HPC); similarly, the ~ 400 configurations analyzed for $SrTiO_3$ via the random sampling approach were calculated in $\sim 0.3 \cdot 10^6$ HPC-core hours. The exhaustive ML-search on a configuration space of millions of polaron pattern requires instead a handful of hours (11 and 2 hours for TiO_2 and $SrTiO_3$, respectively) on a single core of standard personal computers. The bottleneck of the ML-based strategy is the training data generation, which relies on DFT databases; however, our results for $SrTiO_3$ show that the ML model can be efficiently trained by performing a random sampling of the configuration space, relying on few hundreds of polaron energies calculated via DFT. The machine learning workflow proposed in this work is indeed able to overcome the practical limitations of approaches based purely on DFT and on physical intuition. We conclude with a statistical analysis on the polaron patterns explored during the exhaustive search for TiO_2 and $SrTiO_3$ at low defect concentrations, as shown in Supplementary Figure 8. By considering only favorable configurations (*i.e.*, configurations with $E_{\mathrm{pol}}^{\mathrm{ML}}$ at least 80% of the ground state energy), the frequency distribution of polarons in TiO_2 shows a clear tendency towards charge trapping in $\mathrm{Ti}_{\mathrm{S1}}^{\mathrm{A}}$ sites close to the V_{O} at the lowest low defect concentration (Supplementary Figure 8 (a)), correctly capturing the underlying symmetries of the supercell. With increasing defect concentration (see Supplementary Figure 8 (b)), polarons tend to distribute more evenly across all $\mathrm{Ti}_{\mathrm{S1}}^{\mathrm{A}}$ sites compared to the lower concentration, and occupation of $\mathrm{Ti}_{\mathrm{S0}}^{\mathrm{A}}$ sites between the two V_{O} becomes more favorable. In the case of an asymmetric defect pattern, as for $SrTiO_3$, the distribution of favorable sites gets more complicated (see Supplementary Figure 8 (c)). Nevertheless, The ML-aided search identifies sites directly above or below the Nb-dopants in the first two surface layers as the preferred polaron localization sites.

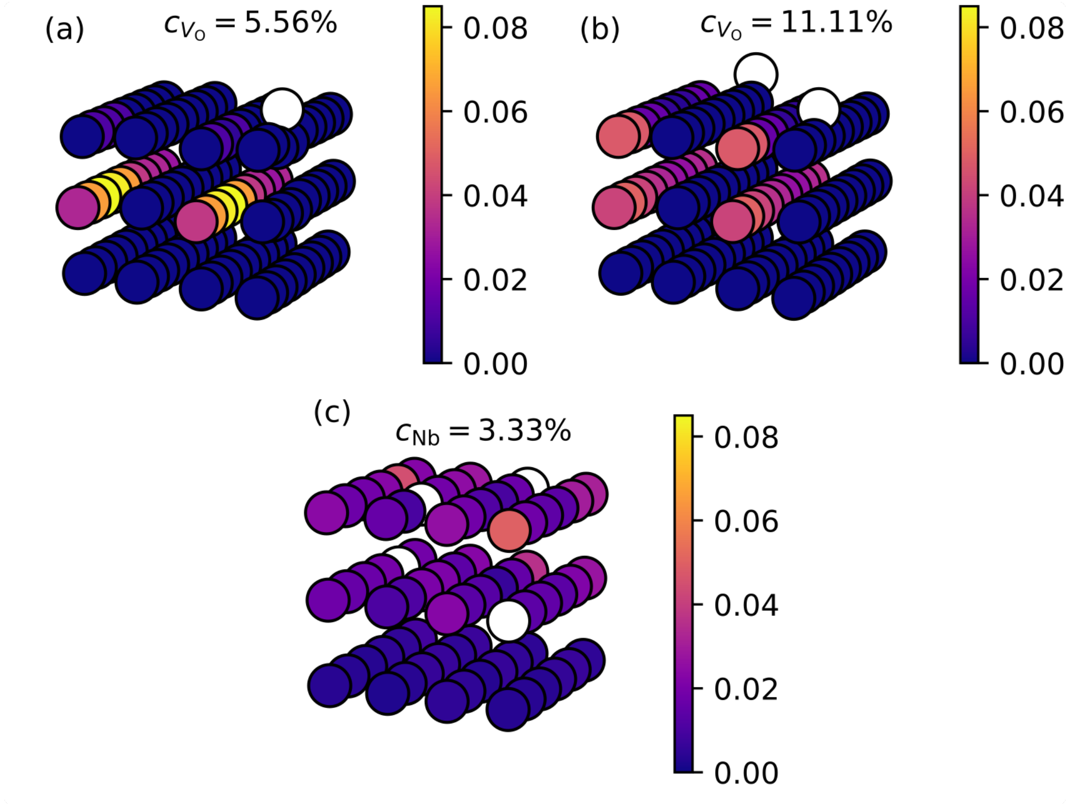
Supplementary Figure 9 shows a statistical analysis of all possible polaron configurations explored using the ML-search, considering polaron formation on Ti_{S0}^{A} and Ti_{S1}^{A} sites in the slab with four polarons (2 oxygen vacancies, $c_{V_{\mathrm{O}}} = 11.1\%$). This results in roughly 53000 total configurations (we have excluded configurations including all the four polarons on the S0 layer). We have collected the polaron formation energies for all configurations, and grouped them depending on the number of polarons per layer. Finally, we have calculated the statistical distribution of the polaron formation energy for every group (including sample mean $\mu(\bar{E}_{\mathrm{pol}}^{\mathrm{ML}})$ and standard deviation $\sigma(\bar{E}_{\mathrm{pol}}^{\mathrm{ML}})$ of the mean polaronic energy, see Supplementary Table 7). As evident from our results, the polaron formation energy depends strongly on the type of hosting site but deviations from the average value are large. These deviations are due to the polaron-polaron and polaron-defect interactions, as we discussed in our recently published study. On top of the primary localization-site dependence of the polaron energy reported in a previous work Deskins et al. (2011), we find a strong influence of the spatial distribution of polarons.



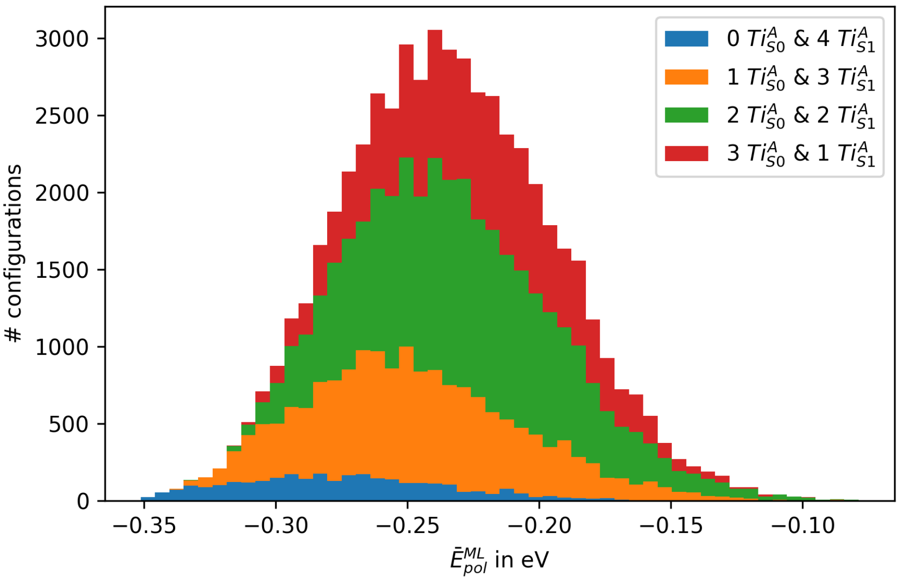
Supplementary Figure 6: Most stable polaron configurations predicted by ML and confirmed at the DFT-level at each c_{V_O} for rutile $TiO_2(110)$. Titanium-sites are shown in blue, polarons are indicated in yellow by a schematic representation of the charge density, and oxygen vacancies are shown in dotted red. For clarity oxygen atoms are not shown and only the three surface layers are displayed.


$c_{\mathrm{Nb}} = 3.33\%$
$c_{\mathrm{Nb}} = 4.17\%$
$c_{\mathrm{Nb}} = 5.0\%$
$c_{\mathrm{Nb}} = 5.83\%$
$c_{\mathrm{Nb}} = 6.67\%$

Supplementary Figure 7: Most stable polaron configurations predicted by ML and confirmed at the DFT-level at each c_{Nb} for $\mathrm{SrTiO_3}(001)$. Titanium-sites are shown in blue, polarons are indicated in yellow by a schematic representation of the charge density, and Nb-dopants are shown in green. For clarity oxygen and strontium atoms are not shown and only three surface layers are displayed.



Supplementary Figure 8: The relative site occupation frequency within the most favorable, ML-searched configurations are displayed for the (a,b) rutile $TiO_2(110)$ and (c) perovskite $SrTiO_3(001)$ surface, respectively. The color encoding and the corresponding lateral colorbars show the percentage of favorable configurations containing a polaron at each Ti-site in the supercell. The positions of defects (V_O and Nb-dopants, respectively) are indicated by a white circle.



Supplementary Figure 9: The mean polaronic energies of various polaron confgurations captured by the ML-model are displayed for $c_{V_{\mathrm{O}}}$=11.1%. Polaron configurations are grouped depending on the number of $\mathrm{Ti}_{\mathrm{S0}}^{\mathrm{A}}$ and $\mathrm{Ti}_{\mathrm{S1}}^{\mathrm{A}}$ polarons (groups are shown in different colours).

Ti_{S0}^{A}	Ti_{S1}^{A}	$\mu(\bar{E}_{\mathrm{pol}}^{\mathrm{ML}})$ (eV)	$\sigma(\bar{E}_{\mathrm{pol}}^{\mathrm{ML}})$ (eV)
0	4	-0.274	0.040
1	3	-0.245	0.040
2	2	-0.229	0.039
3	1	-0.225	0.036

Supplementary Table 7: The mean energies $\mu(\bar{E}_{\mathrm{pol}}^{\mathrm{ML}})$ and the corresponding standard deviations $\sigma(\bar{E}_{\mathrm{pol}}^{\mathrm{ML}})$ of the energy distribution in Supplementary Figure 9 are displayed for possible polaron configuration at $c_{V_{\mathrm{O}}}$= 11.1%. Again, polaron configurations are grouped depending on the number of $\mathrm{Ti}_{\mathrm{S0}}^{\mathrm{A}}$ and $\mathrm{Ti}_{\mathrm{S1}}^{\mathrm{A}}$ polarons in each confguration.